

Universidade de Lisboa

Faculdade de Medicina de Lisboa



**Aplicação de Espectroscopia de FT-IR para a
Serotipagem e Avaliação da Susceptibilidade à
Penicilina em *Streptococcus pneumoniae***

Mónica Cancela de Abreu Gonçalves Vaz Almeida Miranda

Mestrado em Microbiologia Clínica

2006-2008

A impressão desta dissertação foi aprovada pela comissão Coordenadora do Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 26 de Fevereiro de 2008.

Universidade de Lisboa

Faculdade de Medicina de Lisboa



**Aplicação de Espectroscopia de FT-IR para a
Serotipagem e Avaliação da Susceptibilidade à
Penicilina em *Streptococcus pneumoniae***

Mónica Cancela de Abreu Gonçalves Vaz Almeida Miranda

Mestrado em Microbiologia Clínica

Dissertação orientada pelo Professor Doutor João Almeida Lopes

Todas as afirmações efectuadas no presente documento são da exclusiva responsabilidade do seu autor, não cabendo qualquer responsabilidade à Faculdade de Medicina de Lisboa pelos conteúdos nele apresentados.

RESUMO

A espectroscopia de infra-vermelho com transformada de Fourier (FT-IR) aplicada à microbiologia, com o objectivo de estudar microrganismos, tem vindo a ser continuamente desenvolvida nos últimos 10 anos. Esta técnica apresenta vantagens relativamente às técnicas clássicas de biologia molecular, tais como a rapidez e simplicidade de execução assim como a necessidade de poucos consumíveis e um custo reduzido por análise.

O *Streptococcus pneumoniae* é um microrganismo relevante ao nível clínico, sendo fulcral o conhecimento da susceptibilidade ao antibiótico de primeira linha: a penicilina. O estudo deste microrganismo deve incluir a serotipagem, uma vez que se conhecem 91 serótipos de *Streptococcus pneumoniae* e a prevalência de cada serótipo depende de vários factores tais como a zona geográfica e a idade dos indivíduos. São necessárias metodologias rápidas e eficazes para o estudo da susceptibilidade a antibióticos e da serotipagem.

Neste trabalho estudou-se a possibilidade de recorrer à técnica de FT-IR para identificação do serótipo de estirpes de *Streptococcus pneumoniae* e para a determinação da susceptibilidade à penicilina. Foram estudadas 61 estirpes de doentes invasivos, disponibilizadas pelo Laboratório de Microbiologia da Faculdade de Medicina de Lisboa. Estas estirpes foram previamente analisadas quanto à susceptibilidade à penicilina pelo método descrito pelo “Clinical and Laboratory Standards Institute” (CLSI) e quanto ao serótipo, tendo sido identificados os serótipos 9V, 14, 19A, 23F¹. Uma vez que os espectros de FT-IR são de interpretação complexa, recorreu-se à análise

¹ Verificou-se a existência de um grupo de estirpes com pouca expressão de polissacáridos capsulares pequena, que o método de referência não conseguiu identificar

multivariada para o tratamento dos dados. O método FT-IR provou ser eficaz para a classificação segundo o serótipo mas não demonstrou, para as condições estudadas, a capacidade para avaliar a susceptibilidade à penicilina.

Neste trabalho avaliou-se ainda a relação entre a composição em ácidos gordos das estirpes de *S. pneumoniae* e o serótipo assim como a susceptibilidade à penicilina. Nas condições estudadas, não possível demonstrar nenhuma das relações referidas.

Palavras - chave

Streptococcus pneumoniae

Serotipagem

Susceptibilidade à penicilina

Espectroscopia de infra-vermelho com transformada de Fourier (FT-IR)

Ácidos gordos

Análise multivariada

ABSTRACT

In the last decade, Fourier-transform infrared spectroscopy (FT-IR) has been increasingly used as an alternative technique for the identification of microorganisms at the genus, species and subspecies level. FT-IR has the advantage of being fast, simple and a low cost (involving a minor use of reagents and consumables) analytical tool.

Streptococcus pneumoniae is an important microorganism at the clinical level. It is of crucial importance the assessment of the susceptibility to the first line antibiotic: penicillin. The study of this microorganism should include the serotyping as there are 91 different serotypes of *Streptococcus pneumoniae* and the prevalence of each serotype depends on several factors, such as the geographic area and age of individuals. Therefore, it is necessary the existence of rapid and efficient methodologies for the susceptibility and serotyping analysis.

In the present work, FT-IR was proposed as a new method for serotyping *S. pneumonie* and to determine the susceptibility to penicillin. Sixty one invasive strains from the Microbiology Laboratory of the Medicine Faculty of Lisbon were available for the analysis. These strains were previously tested for the susceptibility to penicillin by the method described by the “Clinical and Laboratory Standards Institute” (CLSI) and serotyped according to four different serotypes: 9V, 14, 19A, 23F². The complexity of FT-IR spectra required the application of multivariate analysis in order to unveil spectral features. The method used was effective in the classification according to the serotype but it

² It was identified a group of strains with small expression of capsular polysaccharides for which the reference method was unable to identify the corresponding isolates serotype.

did not prove to be acceptable for the study of penicillin susceptibility, for the adopted conditions.

In the present work the relation between the cells fatty-acids composition with the serotype and the penicillin susceptibility was also evaluated. According to the tested conditions it was not possible to demonstrate any statistically significant relation.

Keywords

Streptococcus pneumoniae

Serotype

Susceptibility to penicillin

Fourier-transform infrared spectroscopy (FT-IR)

Fatty-acids

Multivariate analysis

AGRADECIMENTOS

Este trabalho não estaria concluído se não fosse o apoio que tive no decorrer da parte prática e da escrita desta dissertação.

Queria agradecer, em primeiro lugar, ao INFARMED, I: P., pela disponibilidade de espaço, tempo e consumíveis necessários a toda a parte prática deste trabalho. Do INFARMED, I. P., também não posso deixar de referir o Dr. Luís Soares, impulsionador deste trabalho e sempre disponível para ajudar e compilar bibliografia tão importante para a escrita desta dissertação.

Um agradecimento muito especial para o Professor Doutor João Lopes pela sua extraordinária orientação e incansável disponibilidade para me ensinar a base de quimiometria, essencial para a execução do tratamento de resultados constantes neste trabalho, e para me socorrer nos frequentes momentos de dúvida e desorientação. A distância nunca foi um problema para a resolução de tantas dificuldades que surgiram ao longo do trabalho.

Queria agradecer também ao Professor Doutor Mário Ramirez pela disponibilidade na cedência das estirpes de *Streptococcus pneumoniae* e pelas informações referentes a cada estirpe.

Todo este trabalho não seria possível sem o apoio do meu marido que teve de se ocupar de todos os assuntos da casa e de tomar conta dos nossos filhos muitos fins-de-semana e noites sozinho.

Quero agradecer também a todos os que me ajudaram na elaboração deste trabalho de uma forma directa ou indirecta, tais como a Carlota Dias Pinto e a Ana Galvão.

GLOSSÁRIO

ANN	Rede neuronal artificial	<i>Artificial neural network</i>
ATR	Reflectância Total Atenuada	<i>Attenuated total reflectance</i>
MIC	Concentração mínima inibitória	<i>Minimum inhibitory concentration</i>
DNA	Ácido desoxirribonucleico	<i>Deoxyribonucleic acid</i>
FT-IR	Infra-vermelho com transformada de Fourier	<i>Fourier – transform infrared spectroscopy</i>
GC	Cromatografia gasosa	<i>Gas chromatography</i>
IR	Infra-vermelho	<i>Infra-red</i>
LOBO	Validação cruzada através da eliminação de blocos de exemplos	<i>Leave –one-block-out</i>
LOO	Validação cruzada através da eliminação sequencial de exemplos	<i>Leave –one-out</i>
PBP	Proteínas de ligação à penicilina	<i>Penicillin binding proteins</i>
PCA	Análise de componentes principais	<i>Principal component analysis</i>
PC	Componente principal	<i>Principal component</i>
PCR	Reacção em cadeia da polimerase	<i>Polymerase chain reaction</i>
PLS	Mínimos quadrados parciais	<i>Partial least squares</i>
RMSEC	Raiz do erro quadrado médio da calibração	<i>Root mean square error of calibration</i>
RMSECV	Raiz do erro quadrado médio obtido por validação cruzada	<i>Root mean square error of cross validation</i>
SIMCA	Modelação independente de analogia de classes	<i>Soft independent modelling by class analogy</i>
SNV	Variação padrão normal	<i>Standard normal variate</i>

Os acrónimos utilizados na Tese estão de acordo com o termo correspondente em língua Inglesa.

ÍNDICE

RESUMO.....	I
ABSTRACT	III
AGRADECIMENTOS	V
GLOSSÁRIO	VI
ÍNDICE	VII
OBJECTIVOS.....	IX
I. INTRODUÇÃO	1
1. <i>Streptococcus pneumoniae</i>	1
1.1. Serótipos do <i>Streptococcus pneumoniae</i>	2
1.2. Epidemiologia do <i>Streptococcus pneumoniae</i>	3
1.3. Terapêutica para infecções causadas por <i>Streptococcus pneumoniae</i>	5
2. Espectroscopia de Infra-vermelho com Transformada de Fourier (FT-IR)	10
2.1. FT-IR em Microbiologia	13
3. Cromatografia Gasosa.....	21
4. Quimiometria	22
4.1. Pré-processamento	24
4.1.1. Remoção da média (“mean-centering”)	24
4.1.2. Auto-normalização (“autoscaling”)	25
4.1.3. Derivação (filtro de Savitzky-Golay).....	26
4.1.4. Variação padrão normal (SNV).....	27
4.2. Modelos não supervisionados.....	28
4.2.1. Análise de componentes principais (PCA).....	28
4.2.2. Classificação Hierárquica (HCA)	32
4.3. Modelos Supervisionados	34
4.3.1. Mínimos quadrados parciais (PLS).....	34
4.3.2. Modelação independente de analogia de classes (SIMCA)	36
4.3.3. Redes Neurais Artificiais (ANN).....	37
4.4. Validação dos modelos	40
4.5. Análise dos modelos	41
4.5.1. Estatística de Hotelling (T^2)	41
4.5.2. Estatística dos resíduos (estatística Q)	42
II. MÉTODOS	43
1. Objecto de Estudo	43
1.1. Construção do banco de estirpes e criopreservação	43
1.2. Preparação das estirpes para análise	44
2. FT-IR.....	45
2.1. Análise das amostras por absorvância no suporte de Selenato de zinco	45
2.1.1. Tratamento das estirpes	45
2.1.2. Análise das Suspensões	45
2.1.3. Avaliação da importância da escolha do quadrante de crescimento na análise de FT-IR	47
2.2. Análise das amostras por ATR.....	47
3. Cromatografia Gasosa.....	47
3.1. Preparação das estirpes	48
3.6. Análise por GC.....	50
4. Análise de Resultados	51
4.1. Teste de Qualidade para os espectros de FT-IR	51

4.2. Avaliação dos resultados de GC	52
4.3. Estatística.....	53
4.4. Cálculo da Repetibilidade e da Reprodutibilidade	53
III. RESULTADOS E DISCUSSÃO	55
1. Espectros.....	55
2. Teste de Qualidade	57
3. Cálculo da Repetibilidade e Reprodutibilidade dos Espectros.....	60
4. Verificação da Importância da Escolha dos Quadrantes	63
5. Serotipagem	64
5.1. FT-IR.....	64
5.1.1. PCA	65
5.1.2. SIMCA	83
5.1.3. PLS.....	88
5.1.4. Redes neuronais.....	101
5.2 GC.....	105
6. Susceptibilidade à Penicilina	117
6.1. FT-IR.....	117
6.1.1 PCA	117
6.1.2. SIMCA	119
6.1.3. PLS.....	121
6.1.4. Redes neuronais.....	125
6.2 GC.....	127
7. ATR.....	136
IV. CONCLUSÃO	138
V. PERSPECTIVAS FUTURAS.....	143
VI. BIBLIOGRAFIA.....	145
VII. ANEXOS.....	173
ANEXO I	173
ANEXO II	175
ANEXO III	178

OBJECTIVOS

Esta tese pode ser dividida em duas partes:

1. Desenvolvimento de uma metodologia para a análise de estirpes de *Streptococcus pneumoniae* baseada na espectroscopia de infravermelho com transformada de Fourier (FT-IR). Os resultados obtidos são analisados recorrendo a análise multivariada de modo a verificar se esta será uma abordagem possível para a serotipagem e para a determinação da susceptibilidade à penicilina. Pretende-se também comparar o desempenho dos vários modelos quimiométricos construídos: análise de componentes principais (PCA), modelação independente de analogia de classes (SIMCA), mínimos quadrados parciais (PLS) discriminantes, PLS hierárquico e redes neuronais.
2. Pretende-se ainda estudar a relação da composição em ácidos gordos das estirpes de *Streptococcus pneumoniae* (cadeias de carbono entre 9 e 20 átomos) com o serótipo e com a susceptibilidade à penicilina.

I. INTRODUÇÃO

1. *Streptococcus pneumoniae*

A bactéria *Streptococcus pneumoniae* (*S. pneumoniae* ou pneumococo) foi pela primeira vez isolada por Pasteur e Steinberg há mais de cem anos ⁽¹³¹⁾. É um coco Gram-positivo e encapsulado, sem movimento, não esporulante, é catalase negativo e a temperatura ótima de crescimento é de 37°C. As células com bastante tempo de vida aparentam ser Gram-negativas uma vez que apresentam rápida descoloração quando sujeitas ao teste de Gram. As células têm entre 0,5 a 1,2 µm de diâmetro, são ovais e encontram-se agrupadas aos pares ou em cadeias curtas (Figura I.1).

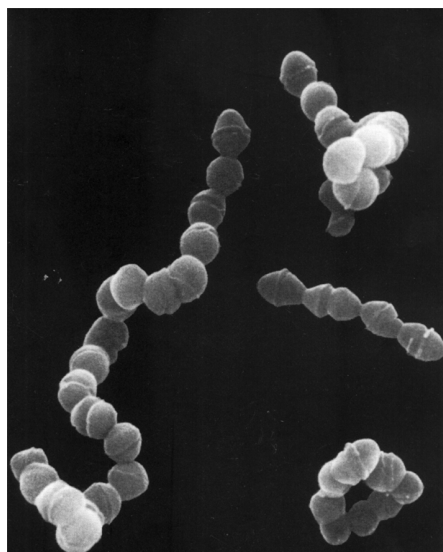


Figura I.1 – Fotografia de *Streptococcus pneumoniae*
(fonte: www.uni-tuebingen.de/modeling/images/strep.gif)

As colónias têm aspectos variados: as que resultam de células encapsuladas são geralmente grandes, redondas e mucóides (1 a 3 mm de diâmetro em gelose de sangue); as colónias de células não encapsuladas são pequenas e achatadas. Todas as colónias passam por um processo de

autólise, ou seja, a parte central das colónias dissolve-se originando uma depressão. As colónias são α -hemolíticas quando o crescimento ocorre à superfície de gelose de sangue em condições aeróbias. Contudo, se forem incubadas em condições anaeróbias, tornam-se β -hemolíticas. A α -hemólise é devida à produção de pneumolisina, uma enzima que degrada a hemoglobina, e que origina o aparecimento de uma cor verde ^(64, 131).

1.1. Serótipos do *Streptococcus pneumoniae*

A cápsula característica dos *S. pneumoniae* é um factor de virulência essencial⁽¹⁷²⁾, pois protege a bactéria da destruição pelos leucócitos⁽¹⁰⁶⁾. Esta cápsula tem uma natureza polissacárida e, consoante a sua composição, assim poderão ser definidos serótipos. Existem 91 serótipos identificados^(104, 120). A técnica clássica para a identificação do serótipo é o teste de Quellung⁽⁶¹⁾ mas uma actualização deste método mais fácil e rápida que se baseia nos mesmos princípios é o “método de xadrez” e foi descrita pela primeira vez em 1993⁽¹⁵⁵⁾. Neste método, numa primeira fase, testam-se soros específicos para um grupo de serótipos. Uma vez conhecido o grupo, determina-se o serótipo da estirpe por reacções de soros específicos de cada serótipo incluído no grupo. Considera-se que um resultado é positivo quando se observa a olho nu, ou com o auxílio de um microscópio de contraste de fase, a presença de uma dilatação da cápsula ou de uma aglutinação⁽¹⁵⁵⁾. Esta técnica tem como desvantagens ser trabalhosa, poder a interpretação do resultado depender da experiência do técnico e necessitar de quantidades de soro elevadas, tornando o método dispendioso⁽⁴²⁾. Actualmente existe um teste “Denka kit” que permite a identificação dos serótipos através de um método de aglutinação.

Relativamente ao teste de Quellung, este teste necessita de apenas um terço do tempo e um terço do custo, oferecendo resultados mais precisos⁽¹⁰³⁾. Outra técnica de serotipagem usada na investigação é a de “dot-blot”^(37, 95). Nesta técnica, são fixados polissacáridos de pneumococos em membranas de nitrocelulose. Em seguida, as membranas são colocadas em contacto com os soros dos serótipos que se pretendem estudar. Um resultado positivo é identificado quando se observam halos escuros nos pontos da membrana correspondentes aos microrganismos com o serótipo testado⁽⁴²⁾. A técnica de reacção em cadeia da polimerase (PCR) também foi utilizada para a identificação de serótipos. Nesta técnica recorre-se a iniciadores de DNA (do inglês *primers*) para sequências específicas de cada serótipo. Como resultado estuda-se a existência, ou a ausência, de amplificação de DNA. Como são conhecidas as posições de cada iniciador, sabe-se qual o serótipo correspondente a cada estirpe testada. É uma técnica rápida, que permite a análise de um vasto número de amostras, e que foi testada nos serótipos mais frequentes em crianças. Quando não se obtém um resultado para os serótipos ensaiados, terá de se recorrer ao “método de xadrez”⁽²⁶⁾.

1.2. Epidemiologia do *Streptococcus pneumoniae*

Os pneumococos estão associados a diversas patologias tais como a pneumonia, otite média, sinusite, endocardite, meningite e até bacteriémia⁽¹³¹⁾. As doenças causadas por *S. pneumoniae* são mais comuns nas crianças. Estima-se que, por ano, cerca de um milhão de crianças morra de doenças pneumocócicas, ocorrendo a maioria dos casos em países em desenvolvimento. Nos países desenvolvidos, os que mais sofrem com as

doenças causadas por *S. pneumoniae* são os idosos e as pessoas pertencentes a grupos de risco tais como os indivíduos infectados pelo vírus de imunodeficiência adquirida (HIV), indivíduos que têm anemias falciformes e quando há falência de órgãos⁽¹⁷²⁾.

A distribuição dos serótipos depende da região geográfica, da idade dos indivíduos e do tipo de infecção, invasiva ou não invasiva. Existem diversos estudos que identificam a predominância dos serótipos segundo os factores apresentados^(58, 100, 144, 145). De acordo com um estudo realizado em crianças saudáveis em Portugal, no período de Janeiro de 2001 a Dezembro de 2003, 79,3% das estirpes pertenciam a um dos serótipos 6B, 14, 23F, 19F e 19A e 13,9% das estirpes estavam associadas aos serótipos 6A, 9V, 33F, 3, 10A, 9A, 15B, 24F e 15A, em que a prevalência de cada um dos serótipos deste último grupo está entre 3,8 a 0,5%⁽¹⁰⁰⁾.

Noutro estudo realizado em Portugal, entre 1999 e 2002, com estirpes de *S. pneumoniae* invasivas de indivíduos com idade igual ou superior a 60 anos (correspondentes a 37% do total das estirpes de pneumococos analisadas) e de idade inferior a 2 anos (correspondentes a 12% do total das estirpes de pneumococos analisadas), verificou-se que a distribuição dos serótipos não é idêntica para os dois grupos etários. Assim, para crianças com menos de 2 anos, os serótipos mais frequentes foram o 14, o 23F, o 6B e o 19A, e para os indivíduos com idade igual ou superior a 60 anos, os serótipos encontrados foram o 3, o 14, o 1, o 8 e o 4⁽¹⁴⁵⁾.

1.3. Terapêutica para infecções causadas por *Streptococcus pneumoniae*

Relativamente à terapêutica de *S. pneumoniae*, o antibiótico de primeira linha é a penicilina. No entanto, as resistências a este antibiótico começam a ter grande significado e encontram-se distribuídas por todo o mundo (como se pode observar na Figura I.2).

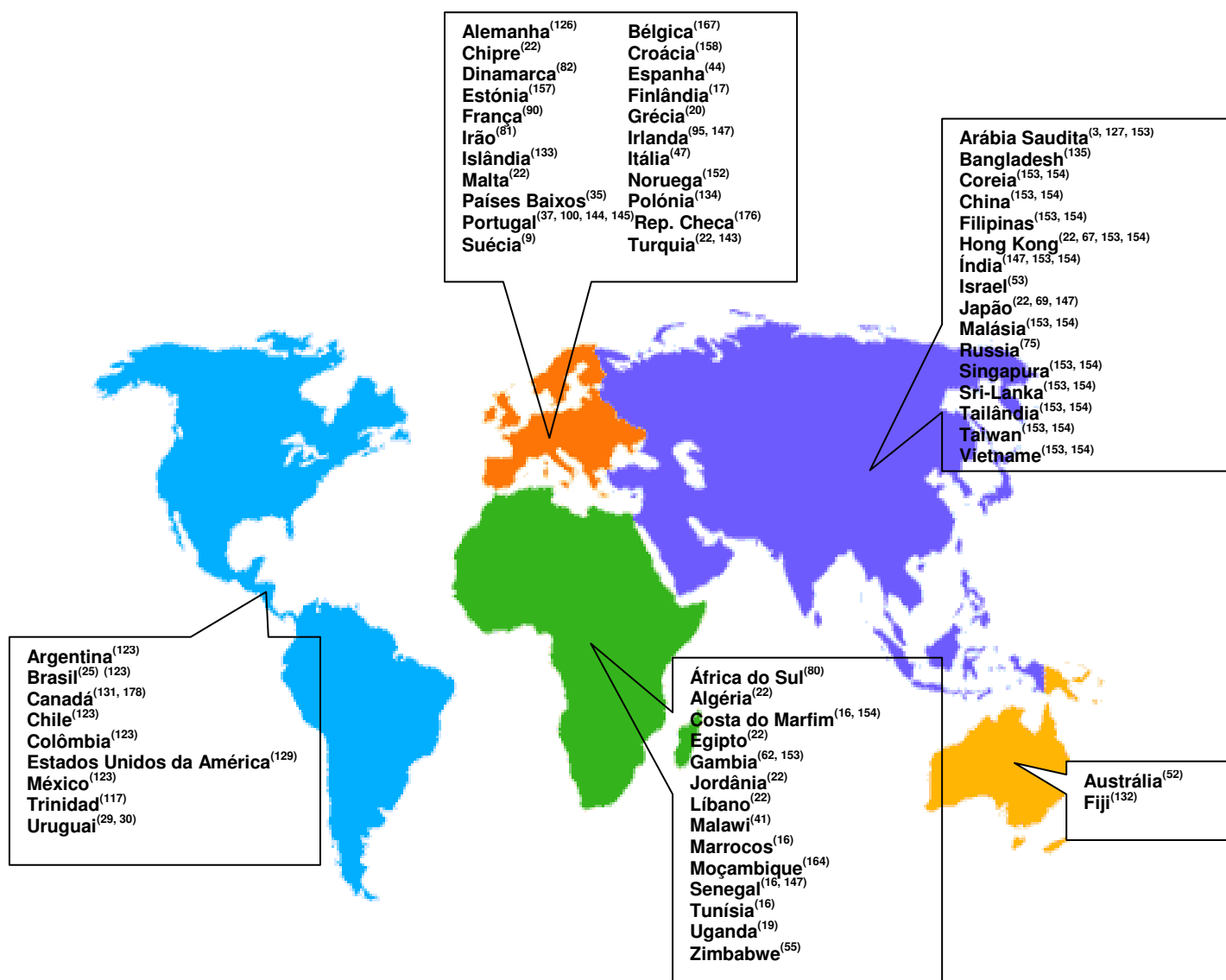


Figura I.2– Distribuição mundial de estirpes de *S. pneumoniae* resistentes à penicilina

Clinicamente, a existência de estirpes resistentes levanta problemas ao nível da terapêutica empírica, uma vez que ainda não existem métodos que permitam o conhecimento do nível de resistência a antibióticos de resultado

imediatos. É importante o conhecimento da susceptibilidade da estirpe causadora da doença ao antibiótico pois, quando se está perante um caso de uma estirpe resistente, existe um agravamento da doença pelo facto de o antibiótico de primeira linha não estar a ser eficaz, sendo necessária a alteração da terapêutica, muitas vezes já tarde^(79, 174).

A penicilina é um antibiótico que pertence à classe dos beta-lactâmicos. Nos pneumococos, a penicilina liga-se às enzimas responsáveis pela síntese da parede celular (PBP), impedindo a sua formação aquando do crescimento e divisão celular. A resistência dos pneumococos à penicilina deve-se a alterações das PBPs que passam a ter uma afinidade mais baixa com o antibiótico⁽⁵⁷⁾. Esta alteração está relacionada com a substituição de genes que representam cerca de 14%⁽¹⁴⁸⁾ a 20%⁽³²⁾ do DNA das PBPs. Outro factor importante para a resistência é a existência de uma enzima (enzima da síntese do muropeptídeo) também alterada, codificada pelo gene *murM*^(36, 79).

É possível definir três níveis de resistência à penicilina: resistente, intermédio ou sensível. No nível intermédio, a estirpe pode ser considerada como sensível ou resistente consoante a história clínica do doente⁽¹⁷⁴⁾. Exceptuando os casos de meningite, quando existe o isolamento de pneumococos considerados de resistência intermédia, a penicilina poderá ser administrada com sucesso^(37, 43, 119, 174).

Os níveis de resistência são avaliados laboratorialmente por métodos de difusão em meio sólido ou o estudo da turbidez de meios líquidos. Os métodos de difusão em meio sólido implicam a incorporação, ou espalhamento, de uma suspensão do microrganismo a testar em meios gelosados. Nestes métodos, a solução de antibiótico a testar pode ser colocada numa cavidade

existente no meio, pode estar em discos de papel que são colocados à superfície do meio sólido ou numa tira de papel impregnada com concentrações crescentes (“*e-test*”). Depois de incubada a placa com o microrganismo a testar e com o antibiótico, é formado um halo à volta do antibiótico – halo de inibição, que será tanto menor quanto maior for a resistência ao antibiótico testado. Assim, para avaliar a susceptibilidade ao antibiótico, para os dois primeiros métodos de difusão em placa, é medido o diâmetro do halo de inibição, e para o método que usa a tira de papel é estudada a concentração a partir da qual se verifica ausência de crescimento. Outra metodologia implica a utilização de meios líquidos. Nestes métodos, é incubada uma solução que contém o microrganismo a estudar e o antibiótico numa dada concentração. São testadas várias concentrações de antibiótico e o resultado é avaliado consoante a presença de crescimento do microrganismo para uma dada concentração de antibiótico. O valor que permite o estudo da susceptibilidade ao antibiótico corresponde à concentração de antibiótico mais baixa que inibiu o crescimento do microrganismo ⁽⁷²⁾.

Existem normas que definem um protocolo para o estudo da susceptibilidade aos antibióticos que especificam, para o procedimento descrito, os valores a partir dos quais se considera uma estirpe sensível, resistente ou de sensibilidade intermédia. O valor de concentração de antibiótico a partir do qual se considera uma estirpe resistente tem a designação de concentração mínima inibitória (MIC) e, para estirpes com os três níveis de susceptibilidade, a MIC é definida por intervalos de concentrações. De acordo com os protocolos definidos pelo “Clinical and Laboratory Standards Institute” (CLSI) para o estudo da susceptibilidade à

penicilina do *S. pneumoniae*, os intervalos são os seguintes: $\leq 0,06$ $\mu\text{g/ml}$ sensível, entre 0,12 e 1 $\mu\text{g/ml}$ considera-se sensibilidade intermédia e ≥ 2 $\mu\text{g/ml}$ resistente⁽³³⁾.

Existem descrições de uma técnica de citometria de fluxo que pretende determinar a MIC dos microrganismos. Esta técnica já foi aplicada, por exemplo, a espécies de *Candida* com o objectivo de avaliar a resistência ao fluconazole⁽⁷³⁾ e ao *Mycobacterium agalactiae* de modo a determinar o MIC para alguns antibióticos⁽⁸⁾. Para avaliar a susceptibilidade à penicilina no *S. pneumoniae* ainda não existem descrições desta técnica. As vantagens da citometria de fluxo são a possibilidade de visualização da heterogeneidade da população no que diz respeito à resistência ao antibiótico, e o tempo para a obtenção de resultados ser menor: os resultados para a determinação da susceptibilidade ao *M. agalactiae* obtêm-se ao fim de 12 h de incubação em vez das 24h previstas nos métodos de referência. A avaliação do resultado é muito rápida (aproximadamente 1 minuto) e não é dependente de uma alteração de cor ou aumento de turbidez, como nos métodos em meio líquido. É, assim, uma boa alternativa aos métodos de difusão em placa ou em meios líquidos descritos anteriormente⁽⁸⁾.

Em Portugal houve uma evolução na taxa de estirpes resistentes nos últimos anos. Em 1989, 4,6% das estirpes de pneumococos eram não susceptíveis à penicilina e, em 1991, este valor subiu para 6,4%. Em 1992 foram identificadas, pela primeira vez, estirpes cuja concentração mínima inibitória para a penicilina foi superior a 2 $\mu\text{g/ml}$. Naquele ano as estirpes resistentes correspondiam a 0,8% da totalidade das estirpes, mas esta percentagem subiu para 5,5% no ano de 1993⁽¹²¹⁾. No período de 1994 a 2004,

a percentagem de estirpes de *S. pneumoniae* invasivas e resistentes foi de 2,6%, enquanto que a percentagem de estirpes com sensibilidade intermédia foi de 18,0%, ou seja, um total de 20,6% de estirpes não susceptíveis⁽³⁷⁾. A não susceptibilidade à penicilina foi associada à não susceptibilidade aos macrólidos, à resistência ao Trimetopim-sulfametoxazole e a uma multirresistência em adultos⁽³⁷⁾. Existem outros estudos, também em pneumococos invasivos, que indicam uma percentagem de estirpes resistentes à penicilina ligeiramente maior (4,7%) e de resistência intermédia idêntica (18,2%)⁽¹⁴⁵⁾.

De acordo com o relatório anual de 2005 da European Antimicrobial Resistance Surveillance System (EARSS), em Portugal, a proporção de estirpes não invasivas de *S. pneumoniae*, não susceptíveis à penicilina, encontra-se entre 25 e 50%⁽³⁹⁾.

Foram descritas associações entre o nível de resistência à penicilina e o serótipo da estirpe⁽¹⁶²⁾. Na Figura I.3 pode-se observar um gráfico, presente no relatório da EARSS, onde se mostra a referida associação⁽³⁹⁾.

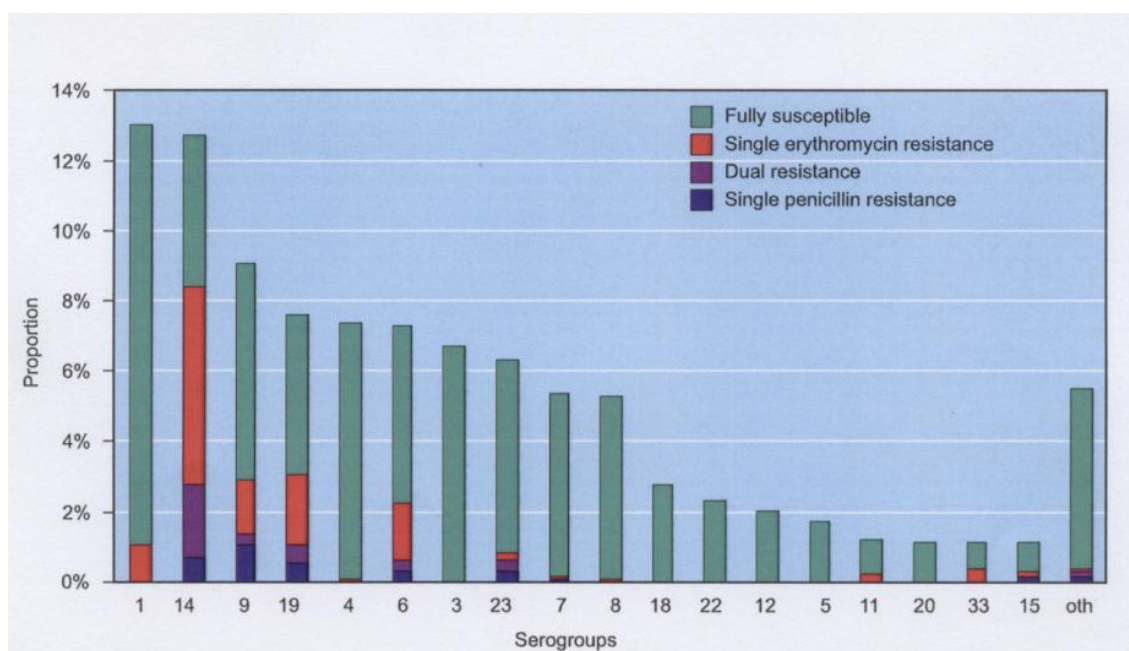


Figura I.3 – Distribuição dos serogrupos e perfil de resistência (penicilina, eritromicina e penicilina+eritromicina), por serogrupo, de estirpes de *S. pneumoniae* constantes na base da EARSS, em 2005. Apenas foram considerados países que forneceram informação de serogrupos para mais de 30 estirpes (Bélgica, República Checa, Dinamarca, Islândia, Eslovénia e Reino Unido) (fonte: relatório da EARSS de 2005⁽³⁹⁾)

Podem ser nomeados outros estudos que relacionam um serótipo com uma dada resistência: a associação do serótipo 14, de estirpes de crianças com idade inferior a 6 anos, à sensibilidade intermédia e resistência à penicilina^(37, 145); a relação entre os serótipos 23F, 6B, 9V, 19A, 19F e 15B, e uma resistência baixa à penicilina, e a multirresistência a vários outros antibióticos⁽¹⁴⁵⁾; e a associação entre os serótipos 14, 9V e 19A e um elevado grau de resistência à penicilina e uma multirresistência para estirpes de indivíduos com mais de 6 anos⁽¹⁴⁵⁾.

2. Espectroscopia de Infra-vermelho com Transformada de Fourier (FT-IR)

A espectroscopia de infra-vermelho estuda a interação da radiação infra-vermelha com a matéria. A região do infra-vermelho pode ser subdividida

em três zonas distintas (NIR, MIR e FIR), de acordo com o número de onda, como mostra a Tabela I.1:

Tabela I.1– Regiões do espectro electromagnético na zona do infra-vermelho⁽¹⁴⁹⁾

Zona do espectro	Número de onda (cm ⁻¹)
Infra-vermelho próximo (NIR)	10-400
Infra-vermelho médio (MIR)	400-4000
Infra-vermelho longínquo (FIR)	4000-14000

Quando a matéria absorve radiação na zona do infra-vermelho, ocorrem alterações do nível de energia vibracional das ligações químicas. Os números de onda em que ocorrem estas transições são característicos de cada grupo funcional⁽³¹⁾. Tendo em conta esta relação, foi possível a elaboração de tabelas que permitem a identificação de moléculas em amostras cuja composição é desconhecida, o que torna a espectroscopia de infra-vermelho muito importante em várias áreas científicas⁽¹⁴⁹⁾.

O que se obtém como resultado de uma análise de espectroscopia de infra-vermelho é um gráfico da intensidade da radiação, em medidas de absorvância ou transmitância, em função do número de onda, designado por “espectro de infra-vermelho”. Este espectro resulta da manipulação de interferogramas de acordo com um algoritmo desenvolvido por Fourier, em que qualquer função matemática pode ser expressa como o somatório de ondas sinusoidais de frequências e amplitudes distintas. A esta técnica dá-se o nome de “espectroscopia de infra-vermelho com transformada de Fourier” (FT-IR)⁽¹⁴⁹⁾.

O primeiro aparelho de FT-IR a ser comercializado foi fabricado pela Digilab, subsidiado pela Block Engineering em Massachussets nos fins dos anos 60. Este primeiro FT-IR incorporava um interferómetro desenvolvido por

Albert Abraham Michelson em 1880, um laser de referência de He-Ne usado para monitorizar a posição do espelho e servir como um padrão do comprimento de onda e um minicomputador. Estes instrumentos permitiam a aquisição de dados com qualidade e com elevada resolução num curto espaço de tempo⁽¹⁴⁹⁾. O interferómetro de Michelson ainda hoje é aplicado na maioria dos espectrómetros de FT-IR. Na Figura I.4 pode ver-se um esquema do interferómetro de Michelson⁽¹⁵⁶⁾.

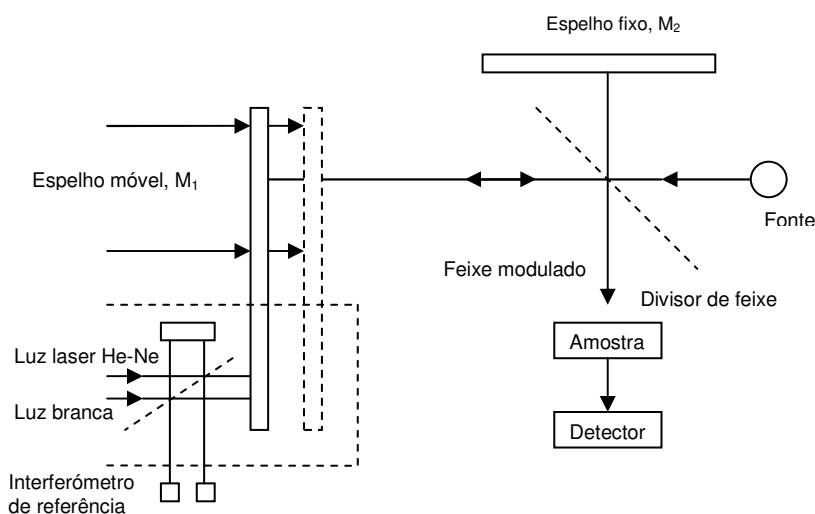


Figura I.4 – Representação esquemática do interferómetro de Michelson (adaptado de ⁽¹⁵⁶⁾)

Os componentes do FT-IR podem ser esquematizados de acordo com a Figura I.5⁽¹⁵⁶⁾.

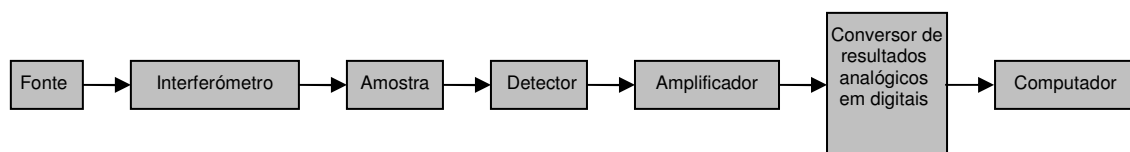


Figura I.5 – Esquema ilustrativo da sequência dos componentes de um FT-IR. (adaptado de ⁽¹⁵⁶⁾).

A escolha do tipo de detector a usar em FT-IR depende da gama espectral em que se pretende trabalhar e da frequência de modulação do interferómetro. Os dois detectores mais usados na zona do infra-vermelho próximo são: (i) o detector piro eléctrico de sulfato de triglicerina deuterada

(DTGS), contido numa janela de halogeneto de metal alcalino, resistente à temperatura, e (ii) o detector de foto condutividade, composto por uma liga semi-condutora de telureto de mercúrio e cádmio (MCT)⁽¹⁴⁹⁾.

Existem diversos modos de amostragem para a análise de microrganismos. Um destes modos recorre ao uso de um suporte de selenato de zinco no qual é introduzida uma suspensão de microrganismos a estudar. Esta suspensão tem de ser imobilizada no suporte, o que é conseguido por aquecimento do suporte numa estufa ou por colocação do suporte num excicador sujeito a vácuo⁽²⁷⁾. Outros modos de amostragem que podem ser referidos são a reflectância total atenuada (ATR), a reflectância difusa (DRIFTS), o GC-FT-IR (acoplamento entre a cromatografia gasosa e o FT-IR), e o micro-FT-IR onde microscópios estão acoplados a aparelhos de FT-IR e que permitem a análise de amostras de diâmetros tão pequenos como 10 μm ⁽¹⁴⁹⁾.

2.1. FT-IR em Microbiologia

Em 1959 foi publicado um artigo onde se refere que as bactérias têm um espectro de infra-vermelho característico, único para cada estirpe, mas devido ao tempo necessário à análise e à metodologia laboriosa de então, conclui-se que não é possível a utilização da espectroscopia de infra-vermelho para a identificação de bactérias⁽¹¹⁵⁾. Em 1962 foram realizados espectros de colónias de *Sphaerotilus natans*, de modo a avaliar a quantidade de um polímero intracelular: ácido poli- β -hidroxibutírico, libertado nas várias fases da curva de crescimento deste microrganismo⁽¹³⁰⁾. As descrições da aplicação da técnica FT-IR à microbiologia, tal como ela hoje se apresenta, devem-se

sobretudo a Naumann e aos seus colaboradores. Nos primeiros trabalhos publicados pelo grupo de Naumann, encontra-se um dendograma que contém várias espécies que se encontram separadas segundo o teste de Gram (positivo ou negativo)^(59, 108-110). Presentemente esta técnica é usada com diversos fins em microbiologia, permitindo já a identificação e caracterização de microrganismos a partir de suspensões preparadas de várias colónias⁽⁹⁸⁾, ou a partir de uma só colónia através da utilização da técnica de microscopia acoplada ao FT-IR^(84,85). Ao contrário dos métodos de identificação bioquímicos, de serotipagem pelo teste de Quellung e de susceptibilidade aos antibióticos, a espectroscopia de FT-IR é uma técnica com resultados rápidos, fácil de executar, com baixo consumo de reagentes e consumíveis, e que necessita de uma concentração celular pequena^(99, 107).

O espectro de infra-vermelho das células microbianas é muito complexo, uma vez que todos os componentes das células contribuem para o espectro final. Naumann elaborou uma tabela onde se encontram discriminados os vários grupos funcionais que podem estar presentes nas células microbianas⁽¹⁰⁷⁾. Geralmente, o estudo de uma determinada característica de uma estirpe recorre a uma ou mais zonas espectrais e não a um só grupo funcional, a um dado número de onda. Uma adaptação da tabela descrita por Naumann pode ser observada na Tabela I.2.

Na Tabela I.2 estão definidas regiões espectrais de acordo com o tipo de moléculas envolvidas: ácidos gordos (w_1 – cor-de-laranja), proteínas (w_2 - verde), DNA, RNA e fosfolípidos (w_3 – cor-de-rosa), polissacáridos da parede celular (w_4 - azul) e região de impressão digital das células

(w₅ - amarelo). Na Figura I.6 estão representadas as zonas espectrais referidas na Tabela I.2.

Tabela I.2 – Relação entre número de onda e grupo funcional existente nas células microbianas (adaptado de⁽¹⁰⁷⁾)

Número de onda	Região espectral	Designação
~3500		Movimentos de distensão O-H de grupos hidroxilo
~3200		Movimentos de distensão N-H das proteínas
~2955	W ₁	Movimentos de distensão assimétricos C-H de -CH ₃ em ácidos gordos
~2930	W ₁	Movimentos de distensão assimétricos C-H de >CH ₂
~2918	W ₁	Movimentos de distensão assimétricos C-H de >CH ₂ em ácidos gordos
~2898	W ₁	Movimentos de distensão C-H em grupos metilo
~2870	W ₁	Movimentos de distensão simétricos C-H de -CH ₃
~2850	W ₁	Movimentos de distensão simétricos C-H de >CH ₂ em ácidos gordos
~1740	W ₂	Movimentos de distensão >C=O dos ésteres
~1715	W ₂	Movimentos de distensão >C=O dos ácidos carbónicos
~1680-1715	W ₂	Movimentos de distensão >C=O dos ácidos nucleicos
~1695	W ₂	Componentes da banda da amida I
~1685	W ₂	Resultante de folhas β – antiparalelas
~1675	W ₂	Configuração β das proteínas
~1655	W ₂	Estruturas α-helicoidais da Amida I
~1637	W ₂	Estruturas das folhas - β da Amida I
~1550-1520	W ₂	Amida II
~1515	W ₂	Banda da tirosina
~1468	W ₃	Deformação C-H de > CH ₂
~1400	W ₃	Movimentos de distensão simétricos C=O de COO ⁻
~1310-1240	W ₃	Banda da amida III das proteínas
~1250-1220	W ₃	Movimentos de distensão assimétricos P=O de >PO ₂ ⁻ de fosfodiésteres
~1200-900	W ₄	Movimentos de distensão C-O e C-C e deformação de C-O-C de hidratos de carbono
~1090-1085	W ₄	Movimentos de distensão simétricos P=O de >PO ₂ ⁻
~720	W ₅	Vibração de rotação da ligação C-H de >CH ₂
~900-600	W ₅	Região da impressão digital do microrganismo

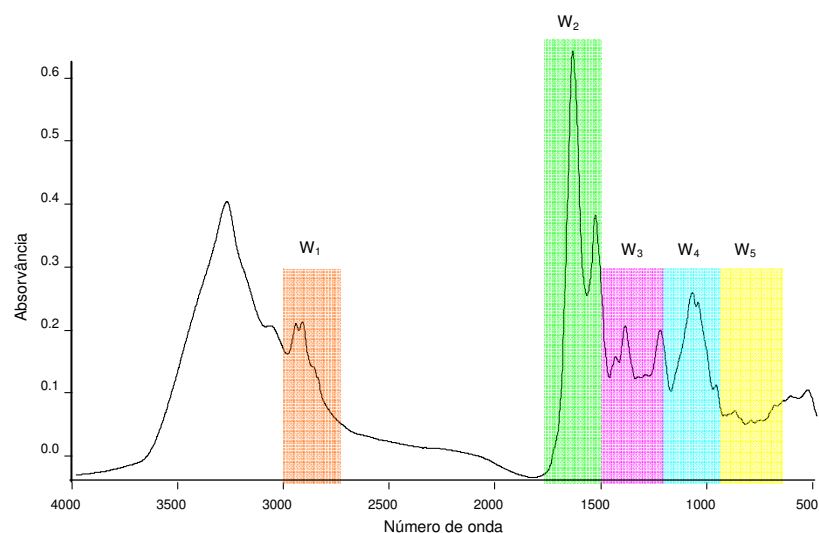


Figura I.6 – Zonas espectrais mais significativas nos espectros de estirpes microbianas, de acordo com o descrito na Tabela I.2.

Existem diversas descrições de aplicações de sucesso em microbiologia, que é bem ilustrada pela listagem das publicações que se encontra na Tabela I.3. Para além das aplicações em microbiologia descritas, é relevante para este trabalho uma publicação que refere um estudo com células cancerígenas em que foi possível distinguir os fenótipos de multi-resistência a medicamentos anticancerígenos⁽⁸⁵⁾.

Na maioria dos trabalhos descritos, os microrganismos são analisados na forma de colónias, dando origem a espectros muito complexos, uma vez que o espectro resultante depende da composição química total das células (do material intracelular, da membrana e da parede). Devido à complexidade do espectro resultante, é indispensável uma avaliação dos espectros por intermédio de análise multivariada (quimiometria)⁽⁸⁹⁾.

Tabela I.3 – Revisão bibliográfica sobre aplicações de FT-IR em microbiologia

Microrganismos	Objectivo do estudo	Referência
<i>Actinomycetes (Maduromysetes, Nocardioforms, Microsporaceae, Streptomycetaceae)</i>	Discriminação de 46 estirpes pertencentes a 4 géneros	(56)
<i>Arthrobacter, Bacillus, Escherichia, Enterobacter, Pseudomonas, Rhodococcus, Serratia, Staphylococcus</i>	Estudo da reprodutibilidade dos espectros ao longo de vários meses	(94)
<i>Bacillus</i>	Descriminação de estirpes pertencentes a 9 espécies representativas de <i>Bacillus</i> spp; e de 38 estirpes, não estirpes de referência, pertencentes a 3 espécies (<i>B. cereus</i> , <i>B. mycoides</i> e <i>B. thurigiensis</i>)	(14)
<i>Bacillus, Pseudomonas, Staphylococcus, Candida</i>	Discriminação de 145 estirpes bacterianas de acordo com o seu género e 24 estirpes classificadas ao nível da espécie; discriminação de <i>C. albicans</i> de acordo com a susceptibilidade a um antibiótico	(163)
<i>Candida albicans, C. dubliniensis, C. stellatoidea</i>	Discriminação de 29 estirpes ao nível da espécie	(160)
<i>Candida albicans, Exophiala dermatitidis, Prototheca</i>	Discriminação de 150 estirpes de <i>Candida</i> , 31 estirpes de <i>Exophiala</i> e 43 estirpes de <i>Prototheca</i>	(139)
<i>Candida dubliniensis</i>	Discriminação entre <i>Candida dubliniensis</i> e <i>Candida albicans</i>	(161)
<i>Clostridium beijerinckii</i>	Discriminação das várias fases de um processo biotecnológico	(141)
<i>Clostridium botulinum</i>	Discriminação de 44 estirpes produtoras de toxina botulínica nos grupos I e II, dependentes das características fenotípicas e genotípicas	(78)
<i>Corynebacterium, Staphylococcus, Enterobacter, Salmonella</i> , entre outros	Avaliação Gram +/-, comparação entre espectros obtidos por reflectância difusa ou transmissão	(88)
<i>Cryptococcus neoformans</i>	Tipagem de 110 estirpes	(93)
<i>Enterobacter, Alcaligenes, Bacillus, Micrococcus, Escherichia, Corynebacterium</i>	Discriminação de Gram+/-; discriminação ao nível da espécie	(21)
<i>Enterococcus faecium</i>	Tipagem de 31 estirpes resistentes à vancomicina - teicoplanina	(122)

Tabela I.3 – Revisão bibliográfica sobre aplicações de FT-IR em microbiologia (Continuação)

Microrganismos	Objectivo do estudo	Referência
<i>Enterococcus faecium</i> , <i>Salmonella enteritidis</i> , <i>Bacillus cereus</i> , <i>Yersinia enterocolitica</i> , <i>Shigella boydii</i> , <i>Escherichia coli</i>	Quantificação de bactérias em alimentos	(54)
<i>Escherichia coli</i>	Estudo estrutural da superfície de estirpes de <i>E. coli</i> O123 pertencentes ao subtipo de antígenos I ou II	(18)
<i>Escherichia coli</i>	Descriminação entre estirpes susceptíveis a beta-lactâmicos e transconjugantes resistentes	(24)
<i>Escherichia coli</i> , <i>Salmonella typhimurium</i>	Identificação em alimentos	(28)
<i>Escherichia coli</i> , <i>Micrococcus luteus</i> , <i>Lactobacillus plantarum</i> , <i>Pseudomonas aeruginosa</i> , <i>Geobacillus caldxylosilyticus</i> , <i>Thermobrachium celere</i> , entre outras	Caracterização e diferenciação entre estirpes mesófilas e termófilas	(45)
<i>Escherichia coli</i> , <i>Proteus mirabilis</i> , <i>Klebsiella</i> , <i>Pseudomonas aeruginosa</i> , <i>Enterococcus</i>	Discriminação de 59 estirpes pertencentes a 5 géneros	(50)
<i>Escherichia coli</i>	Estudo da influência de várias concentrações de penicilina	(177)
<i>Lactobacillus</i>	Identificação de 42 estirpes de <i>Lactobacillus</i> extraídas de grãos de Kefir e de 12 estirpes de referência	(23)
<i>Lactobacillus</i>	Discriminação e classificação de 53 estirpes pertencentes a 13 espécies diferentes	(34)
<i>Streptococcus</i> , <i>Lactococcus</i> , <i>Lactobacillus</i> , <i>Leuconostoc</i> , e <i>Pediococcus</i>	Identificação de 22 estirpes identificadas ao nível do género e 38 estirpes de <i>Lactobacillus</i> identificadas ao nível da espécie	(38)
<i>Lactococcus lactis</i>	Discriminação de 7 estirpes pertencentes a 2 subespécies	(92)
<i>Listeria</i>	Discriminação de 7 estirpes pertencentes a 7 espécies; estudo de 5 estirpes pertencentes a 5 serótipos que não conseguem ser discriminadas ao nível do serótipo mas ao nível de 2 grupos	(91)
<i>Listeria</i>	Discriminação de 7 estirpes pertencentes a 7 espécies	(63)
<i>Listeria</i>	Discriminação de 24 estirpes de <i>L. monocytogenes</i> de acordo com o serótipo	(91)
<i>Listeria monocytogenes</i>	Discriminação de 12 serovares de 106 estirpes. Validação com 166 estirpes	(125)

Tabela I.3 – Revisão bibliográfica sobre aplicações de FT-IR em microbiologia (Continuação)

Microrganismos	Objectivo do estudo	Referência
<i>Legionella spp.</i>	Discriminação de 10 estirpes em 3 serogrupos	(66)
<i>Microsistis, Anabaena, Chlorella, Synechoccus, Oscillatoria</i>	Discriminação de 6 estirpes	(74)
<i>Mycobacterium sp.</i> que não provoca tuberculose	Identificação de 28 estirpes	(124)
<i>Mycobacterium bovis</i>	Tipagem de 95 estirpes de 10 spoligotipos	(169)
<i>Pseudomonas</i>	Discriminação de 240 estirpes pertencentes a 15 espécies	(71)
<i>Pseudomonas</i>	Estudo da influência do mercúrio em 450 estirpes de <i>Pseudomonas</i> recuperadas de solos	(65)
<i>Pseudomonas, Bacillus, Candida</i>	Identificação ao nível da espécie e estudo da susceptibilidade para a <i>Candida albicans</i>	(140)
<i>Pseudomonas aeruginosa, Escherichia coli</i>	Identificação em águas	(2)
<i>Pseudomonas aeruginosa</i>	Estudo da susceptibilidade ao imipenemo	(150)
<i>Pseudomonas aeruginosa, Escherichia coli</i>	Discriminação de acordo com a susceptibilidade a um antibiótico	(151)
<i>Pseudomonas aeruginosa, Pseudomonas fluorescens, Pseudomonas putida, Pseudomonas stutzeri</i>	Estudo para a identificação em amostras de solos	(175)
<i>Saccharomyces, Candida, Clavispora, Dekkera, Debaryomyces</i> , entre outros	Classificação de 332 estirpes pertencentes a várias espécies e géneros	(86)
<i>Saccharomyces cerevisiae</i>	Discriminação de 22 estirpes ao nível da estirpe	(159)
<i>Sacharomyces cerevisiae/ Hanseniaspora uvarum e Lactobacillus acidophilus/ S. salivarius ssp. thermophilus</i>	Quantificação da relação entre dois microrganismos em populações mistas de iogurte	(116)
<i>Salmonella enterica</i>	Discriminação de vários serovares	(10)
<i>Salmonella enterica</i>	Estudo da influência do meio de crescimento na discriminação de serovares	(11)
<i>Salmonella enterica</i>	Serotipagem através de extractos lipopolissacáridos	(76)

Tabela I.3 – Revisão bibliográfica sobre aplicações de FT-IR em microbiologia (Continuação)

Microrganismos	Objectivo do estudo	Referência
<i>Salmonella enterica</i>	Serotipagem através de proteínas da membrana	(77)
<i>Salmonella enteridis</i>	Classificação de estirpes clínicas, comparação com outros métodos	(142)
<i>Serratia marcescens</i>	Classificação de 66 estirpes clínicas	(68)
<i>Staphylococcus aureus</i>	Discriminação de 85 estirpes pertencentes a subtipos de MRSA	(4)
<i>Staphylococcus aureus</i>	Discriminação de estirpes nativas de variantes com colónias pequenas	(15)
<i>Staphylococcus aureus</i>	Estudo da susceptibilidade à metilina de 37 estirpes (32 estirpes susceptíveis e 15 resistentes)	(51)
<i>Staphylococcus aureus</i> e <i>Staphylococcus sp</i>	Discriminação de 39 estirpes de várias espécies e subespécies de <i>Staphylococcus</i> presentes no leite e queijo	(87)
<i>Staphylococcus epidermidis</i>	Caracterização da composição de estirpes de <i>S. epidermidis</i> , ao nível molecular	(5)
<i>Streptococcus</i> , <i>Lactobacillus</i>	Estudos taxonómicos de 2 espécies de <i>Streptococcus</i> e 3 de <i>Lactobacillus</i>	(6)
<i>Streptococcus</i> , <i>Lactobacillus</i> , <i>Lactococcus</i> , <i>Leuconostoc</i>	Identificação ao nível do género, espécie e subespécie	(7)
<i>Streptococcus</i> , <i>Enterococcus</i>	Discriminação de 19 estirpes pertencentes a 6 espécies de 2 géneros	(49)
<i>Streptococcus</i> , <i>Staphylococcus</i> , <i>Clostridium</i> , <i>Aeromonas</i> , <i>Citrobacter</i> , <i>Enterobacter</i> , entre outros	Validação de uma biblioteca	(59)
<i>Streptococcus</i> , <i>Staphylococcus</i> , <i>Clostridium</i> , <i>Legionella</i> , <i>Escherichia coli</i>	Classificação de 139 estirpes pertencentes a 14 espécies (3 géneros); discriminação de Gram +/-	(60)
<i>Streptococcus</i>	Discriminação ao nível da espécie de 40 estirpes pertencentes a 9 espécies e separação das estirpes em 2 grupos: mutantes de <i>Streptococcus</i> e outros <i>Streptococcus</i>	(165)
<i>Streptococcus mitis</i>	Estudo da influência do meio de crescimento	(166)
<i>Streptococcus pneumoniae</i>	Discriminação de estirpes consoante serótipo e quantidade de polissacáridos das cápsulas responsáveis pelos fenótipos opacos / transparentes	(136)
<i>Trychophyton</i> , <i>Microsporum</i>	Discriminação de 27 estirpes pertencentes a 3 espécies (2 géneros)	(13)

3. Cromatografia Gasosa

No campo da microbiologia foi desenvolvida uma técnica de identificação de microrganismos com base na identificação dos ácidos gordos constituintes das células. A técnica mais adequada para a identificação de ácidos gordos é a cromatografia gasosa⁽¹¹²⁾.

As membranas celulares dos microrganismos contêm ácidos gordos. Consoante o género e a espécie do microrganismo, assim será o tipo e a quantidade de ácidos gordos. A composição em ácidos gordos da membrana de um dado microrganismo também depende das condições ambientais e de crescimento. Uma vez conhecida a composição em ácidos gordos das membranas de vários microrganismos, todos crescidos nas mesmas condições, pode-se construir uma biblioteca de cromatogramas para futuras comparações. Este processo foi desenvolvido pela MIDI e o procedimento tem o nome de Sherlock^{®(1)}. Neste procedimento, os ésteres metílicos dos ácidos gordos da membrana, com mais de 9 átomos de carbono e menos de 20, são analisados por cromatografia gasosa e o perfil obtido para o microrganismo a identificar é comparado com os perfis de microrganismos existentes numa base de dados no computador. Quanto menor for a distância multivariada entre o perfil dos ácidos gordos do microrganismo a identificar e o perfil de uma espécie conhecida e existente na base de dados, maior será o índice de semelhança (SI). Uma coincidência exacta a nível qualitativo e quantitativo entre o microrganismo a identificar e um existente na base de dados terá um índice de semelhança de 1,000⁽¹⁾.

Quando não são seguidas as indicações da MIDI para o cultivo dos microrganismos, não se deve aceitar a identificação da espécie e o índice de

semelhança dada pelo programa Sherlock[®] como verdadeira, sem recorrer a outros testes que o confirmem. No entanto, o cromatograma obtido depende do microrganismo, das suas condições de cultivo e das condições de análise definidas para a cromatografia gasosa e não de manipulação do programa Sherlock[®]. Assim, poder-se-á usar este método para gerar cromatogramas que permitirão comparar estirpes que tenham sido cultivadas nas mesmas condições⁽¹⁾.

4. Quimiometria

A designação “quimiometria” foi introduzida em 1972 pelo sueco Svante Wold e pelo americano Bruce R. Kowalski⁽¹¹⁸⁾. No entanto, foi apenas em 1974, aquando da fundação da International Chemometrics Society, que surgiu a primeira descrição desta disciplina. “Quimiometria”⁽⁴⁰⁾ pode ser definida como a disciplina química que usa métodos matemáticos, estatísticos e lógica formal para:

- (a) desenhar e seleccionar experiências e procedimentos de medida ideais;
- (b) extrair o máximo de informação química relevante dos resultados químicos obtidos;
- (c) conhecer os sistemas químicos através dos métodos matemáticos e estatísticos que foram inicialmente aplicados à psicologia, biologia e agricultura no início do século passado.

Com as ferramentas da quimiometria é possível a detecção de pequenas alterações no objecto de estudo que não se conseguem distinguir apenas por observação visual dos espectros. Os espectros de FT-IR de

estirpes microbiológicas são muito complexos e, sem utilização de ferramentas estatísticas, não são de grande utilidade. No entanto, existem estudos publicados em que, apenas por sobreposição de espectros, foi possível distinguir estirpes, tais como: estirpes Gram-positivas de estirpes Gram-negativas^(60, 109, 110), estirpes de *Escherichia coli* crescidas na presença ou ausência de antibióticos⁽¹⁷⁷⁾ e estirpes de *Streptococcus mitis* crescidas em dois meios distintos⁽¹⁶⁶⁾.

Na análise dos espectros FT-IR de microrganismos, são usadas duas abordagens quimiométricas. A primeira baseia-se em métodos não supervisionados através dos quais são analisados os dados, independentemente da identidade da amostra. Estes são considerados métodos qualitativos, uma vez que o seu objectivo principal é o de distinguir objectos ou populações. Exemplo destes métodos é a análise de componentes principais (PCA) e a classificação hierárquica. A segunda abordagem baseia-se em métodos supervisionados, métodos em que é necessário o conhecimento da identidade da amostra. Nesta última abordagem, parte-se de um grupo de amostras, bem caracterizadas, que serão usadas na construção do modelo, que será depois usado para a determinação da identidade de uma amostra desconhecida. Exemplos destes métodos supervisionados são a modelação independente de analogia de classes (SIMCA), os mínimos quadrados parciais (PLS) e as redes neurais⁽⁴⁸⁾.

O método escolhido deverá ser adequado ao objectivo da análise. Idealmente o método deverá poder ser interpretado do ponto de vista biológico⁽⁴⁸⁾.

Em seguida, apresentam-se várias ferramentas de quimiometria habitualmente usadas na avaliação de espectros de FT-IR e exemplos de publicações em que se refere a sua aplicação.

4.1. Pré-processamento

O pré-processamento é a aplicação de um ou vários métodos correctivos de interferências causadoras de variação, na aquisição de espectros. Os espectros de FT-IR são muito complexos e contêm muita informação ao nível químico e físico. Variações de temperatura, tamanho da partícula e quantidade de amostra influenciam os espectros. Este tipo de variações, assim como o efeito causado pela dispersão da luz, deve ser removido de modo a evidenciar apenas as informações químicas. Com as ferramentas de quimiometria, pretende-se construir um modelo com base nas características das estirpes analisadas e conhecidas *a priori* e, para isso, deve-se escolher o método de pré-processamento mais ajustado. O pré-tratamento dos dados, em regra, melhora o modelo, embora seja necessário entender o tipo de tratamento utilizado de forma a ser usado da melhor maneira⁽¹⁰¹⁾. Descrevem-se, de seguida, os tipos de pré-processamento usualmente utilizados em espectroscopia.

4.1.1. Remoção da média (“mean-centering”)

O objectivo da remoção da média é eliminar o valor da intensidade de cada variável, revelando apenas as flutuações dos dados em redor do valor médio.

Nesta função, cada variável, x_{ik} , é centrada (x_{ik}^*) através da subtracção da resposta de cada variável sobre a resposta média dessa variável nos dados, em todas as amostras. A operação de remoção da média pode ser representada pela Equação I.1.

$$x_{ik}^* = x_{ik} - \bar{x}_k \quad \text{Equação I.1}$$

Os dados a serem tratados encontram-se na forma de uma matriz, em que cada coluna corresponde a um número de onda e cada linha a uma amostra. Assim, i representa o índice da linha da matriz (para as matrizes usadas neste trabalho, i corresponde a uma amostra ou, mais correctamente, a um espectro), e k o índice da coluna (para as matrizes usadas neste trabalho, k corresponde a um número de onda), \bar{x}_k representa a média da coluna k da matriz que é calculada de acordo com a equação seguinte:

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik} \quad \text{Equação I.2}$$

Nesta equação, n representa o número de linhas da matriz usadas para o cálculo da média⁽¹¹⁸⁾.

4.1.2. Auto-normalização (“autoscaling”)

A auto-normalização é usada quando há uma diferença muito grande entre colunas (variáveis), devido a valores absolutos muito diferentes ou gamas distintas, o que acontece quando os dados resultam de diferentes tipos de instrumentos ou quando as unidades das medidas não são iguais para todas as variáveis. Nestes casos, se não se realizar a auto-normalização, as variáveis com a amplitude maior vão tender a dominar o processo de modelação e aquelas com amplitude menor vão tender a ser ignoradas.

Na auto-normalização obtém-se um valor para cada variável (x_{ik}^*) que resulta da remoção da média, tal como descrito no ponto anterior, e de uma divisão pelo desvio padrão das variáveis:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{S_k} \quad \text{Equação I.3}$$

Aqui, x_{ik} representa a variável que está na linha i e coluna k da matriz, \bar{x}_k a média da coluna k e S_k representa o desvio padrão das variáveis e é calculado de acordo com a Equação I.4:

$$S_k = \sqrt{\frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{(n-1)}} \quad \text{Equação I.4}$$

Nesta equação, n representa o número de objectos.

Os dados normalizados têm a característica de que cada variável tem uma média de 0 e um desvio padrão de 1⁽¹¹⁸⁾.

Amiali e seus colaboradores utilizaram este pré-processamento para analisar espectros de FT-IR de *Staphylococcus aureus*⁽⁵⁾.

4.1.3. Derivação (filtro de Savitzky-Golay)

Os filtros de Savitzky-Golay foram, pela primeira vez, descritos em 1964, com o objectivo de serem aplicados a dados que expressassem uma propriedade física contínua, de modo a facilitar a compreensão dos mesmos⁽¹³⁸⁾. No caso dos espectros de FT-IR, a propriedade física contínua é a absorvância ou a transmitância.

Nos filtros de Savitzky-Golay os sinais são ponderados por coeficientes. Para se determinarem os coeficientes apropriados, é necessário

decidir a ordem do filtro (máximo expoente) e o tamanho do filtro (dimensão da janela). De seguida, obtêm-se os coeficientes c_j a partir dos valores tabelados e dividem-se por uma constante - NORM (dependente da ordem e do tamanho da janela do filtro)⁽¹¹⁸⁾:

$$y_i = \frac{1}{NORM} \sum_{j=-m}^{j=m} c_j y_{k+j} \quad \text{Equação I.5}$$

Estes filtros podem ser usados para a derivação de sinais em forma de curvas através de uma fórmula discreta onde é necessário colocar coeficientes apropriados que se encontram tabelados. Esta derivação permite a remoção do ruído de fundo de um sinal, a determinação da posição de um pico e a resolução de picos a nível visual. Em aplicações espectroscópicas, a primeira derivada remove as variações aditivas da linha de base do espectro nos perfis espectrais⁽¹¹⁸⁾.

Na parte prática deste trabalho define-se a função de derivação de Savitzky-Golay como “SAVGOL” e os parâmetros do filtro como (a,b,c), sendo \underline{a} o tamanho do filtro, \underline{b} a ordem do polinómio e \underline{c} a ordem da derivada.

Existem diversos artigos que descrevem a utilização deste pré-processamento^(5, 15, 23, 24, 49, 65, 93, 113, 125, 161, 169).

4.1.4. Variação padrão normal (SNV)

Este método tem como objectivo centrar e ajustar a escala dos espectros individualmente. O método é descrito pela Equação I.6.

$$\hat{X}_{ik} = \frac{(x_{ik} - m_i)}{s_i} \quad \text{Equação I.6}$$

Nesta equação, x_{ik} é o valor da absorvância para o número de onda k relativo ao espectro i , m_i é a média de cada espectro i e S_i é o desvio padrão das absorvâncias relativas ao espectro i . A operação SNV realiza um ajuste aditivo e multiplicativo em que cada espectro é normalizado com base nos dados apenas daquele espectro, ou seja, não depende dos outros espectros a analisar. Para cada espectro da amostra, o desvio ajustado é simplesmente a média de todos os valores sobre todas as variáveis e o ajuste multiplicativo é o desvio padrão dos valores de todas as variáveis⁽¹⁰⁵⁾.

Existem outras formas de pré-processamento espectral que não foram consideradas neste trabalho, pelo que não são descritas nesta secção (e.g., correcção multiplicativa de sinal).

4.2. Modelos não supervisionados.

4.2.1. Análise de componentes principais (PCA)

A análise de componentes principais (PCA) é uma ferramenta muito utilizada em quimiometria e bastante aplicada na análise de espectros de FT-IR^(2, 5, 24, 49, 63, 85, 87, 111, 150, 151, 159).

O método PCA é um processo matemático que transforma um número de variáveis correlacionadas num número mais pequeno de variáveis não correlacionadas que são chamadas “componentes principais” (PCs) e que representam as variações principais do espectro. O primeiro componente principal (PC) é a combinação de variáveis que explica o máximo possível da variância, o segundo PC é a combinação de variáveis que explica o máximo possível da variância ainda não explicada pelo primeiro PC, e que é

independente (ortogonal) do primeiro. O último PC será o que menor contribuição dá para a explicação da variância total dos dados originais⁽²⁾.

Na análise de componentes principais é possível detectar a existência de espectros aberrantes através da estatística dos resíduos (Q) e da estatística de Hotelling, como se descreve no ponto 4.5 deste capítulo.

Matematicamente, cada componente principal é uma combinação linear das variáveis originais (ver Equação I.7) ⁽¹²⁸⁾.

$$T_j = p_{1j}X_1 + p_{2j}X_2 + \dots + p_{ij}X_i \quad \textbf{Equação I.7}$$

Geometricamente os componentes principais T_j podem ser vistos como projecções dos dados originais sobre eixos ortogonais (pesos ou “loadings”, p_i) que cobrem o espaço das variáveis X .

Os pesos permitem interpretar os componentes principais em função das variáveis com os quais se encontram mais correlacionados. Da soma dos quadrados dos pesos das variáveis para cada factor resulta o valor próprio da componente⁽¹²⁸⁾. O grau de correlação entre as variáveis é proporcional ao cosseno do menor ângulo entre os pesos de quaisquer duas variáveis (considerando apenas os primeiros componentes principais). Assim, num mesmo mapa de pesos, é possível visualizar em simultâneo a correlação entre todas as variáveis (se for garantido que esse mapa contabiliza uma quantidade de variância suficiente). Pontos que estejam próximos indicam forte correlação positiva, enquanto que pontos diagonalmente opostos indicam forte correlação negativa⁽⁹⁶⁾.

Em termos matriciais, o modelo PCA pode ser escrito através da Equação I.8.

$$X = T \times P^t + E \quad \text{Equação I.8}$$

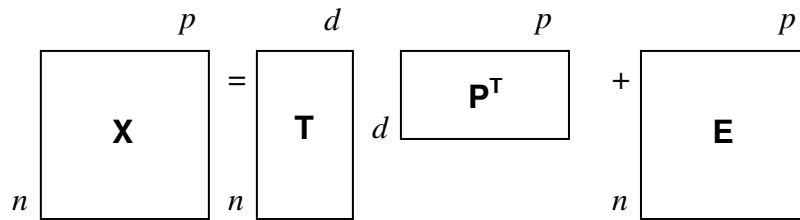


Figura I.7 – Representação gráfica da operação de decomposição da matriz X através da técnica de análise de componentes principais⁽⁹⁶⁾.

Se se considerar um conjunto de dados espectrais, então a matriz X contém os espectros, a matriz T os componentes principais, a matriz P os pesos e a matriz E os resíduos.

Para uma correcta análise de PCA deve definir-se o número de componentes principais adequado. Para isso não existe nenhuma regra teórica, mas existem regras empíricas que serão apresentadas em seguida⁽¹¹⁸⁾:

1. Percentagem de variância explicada: cada PC está associado a uma variância e a representação gráfica de um PC em função de outro PC tem uma variância específica que resulta da soma da variância explicada por cada PC. Pode haver um critério de variância explicada mínima num dado trabalho.
2. Critério dos valores próprios: existe uma regra segundo o qual os valores próprios não devem ser inferiores a 1. Então, o número máximo de componentes principais a usar deverá ser aquele que tenha os valores próprios superiores a 1.
3. Representação gráfica dos valores próprios em função do número de componentes: nesta representação gráfica há uma regra segundo a qual

que o número de componentes principais a usar deverá ser aquele imediatamente anterior à diminuição drástica do declive.

4. Validação cruzada⁽¹⁰¹⁾: esta é a forma mais correcta de determinação do número de PCs adequado. De uma forma simplificada, o procedimento é o seguinte: um espectro é retirado do conjunto total de espectros e é calculado o modelo PCA com os restantes espectros. Depois, esse espectro é testado no modelo construído, obtendo-se uma previsão. Essa previsão é comparada com o valor experimental correspondente a essa amostra, obtendo-se um erro de previsão (residual). Em seguida, é retirado outro espectro e colocado de novo o primeiro. O modelo é calculado novamente e este espectro é testado no segundo modelo construído. Fazendo este exercício para todos os espectros, pode calcular-se a soma da raiz quadrada dos residuais. O número de componentes principais adequado será aquele que apresenta um erro residual mais baixo: RMSECV (“Root Mean Square Error of Cross Validation”), que pode ser determinado pela Equação I.9. Este método tem o nome de validação cruzada do tipo “leave-one-out” (LOO).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad \text{Equação I.9}$$

Na equação I.9, y_i representa o valor experimental da variável dependente relativa a cada espectro i , \hat{y}_i representa o valor previsto pelo modelo (é importante salientar que cada \hat{y}_i é obtido por um modelo distinto gerado com as restantes N-1 amostras) e N representa o número total de espectros.

Pode também fazer-se este tipo de validação cruzada de outras formas, por exemplo, em vez de se retirar um espectro de cada vez, retira-se um conjunto de espectros, neste caso a técnica designa-se por “leave-one-block-out” (LOBO).

4.2.2. Classificação Hierárquica (HCA)

Neste método, os objectos são agregados passo a passo de acordo com a semelhança das suas características. Como resultado, obtém-se uma representação gráfica baseada em distâncias entre espectros. Nestes gráficos, duas amostras são mais semelhantes quanto menor for a distância entre elas. Estas distâncias, na maioria das vezes, são as distâncias euclidianas ou de Manhattan.

A distância euclidiana entre dois espectros (x_i e x_j) é definida pela Equação I.10:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad \text{Equação I.10}$$

A distância de Manhattan, d_{ij} , entre dois espectros (x_i e x_j) é definida pela Equação I.11:

$$d_{ij} = \|x_i - x_j\| \quad \text{Equação I.11}$$

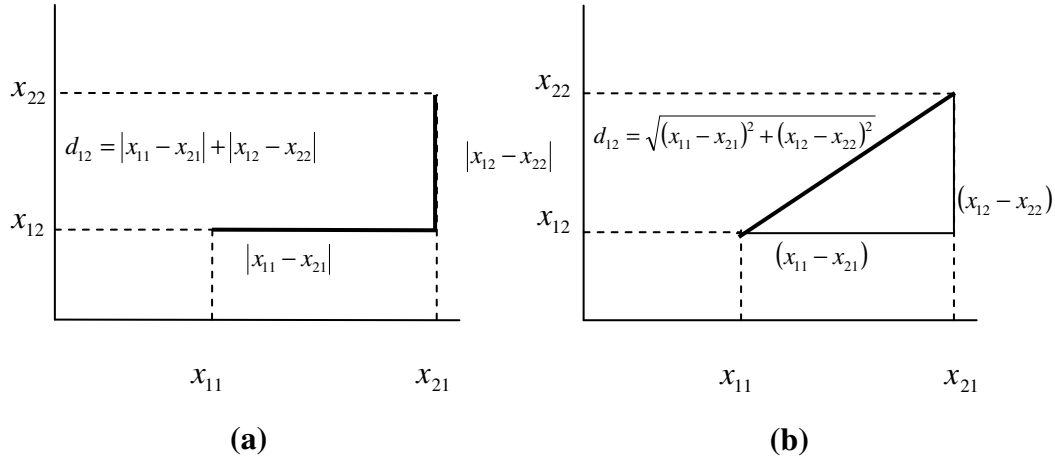


Figura 1.8 – Representação esquemática das distâncias de Manhattan (a) e euclidiana (b)

Existem outros dois tipos de distâncias: a de Mahalanobis e a de Pearson. Na distância de Mahalanobis são tidas em conta diferentes escalas das variáveis e a sua correlação. A fórmula de cálculo é descrita pela Equação 1.12:

$$D_{ij} = (x_i - x_j)^t C^{-1} (x_i - x_j) \quad \text{Equação 1.12}$$

Nesta equação, C é a matriz de co-variância e x_i e x_j correspondem aos vectores das colunas para os objectos i e j , respectivamente. A aplicação desta distância é referida, por exemplo, no estudo de *Lysteria monocytogenes*⁽⁹¹⁾.

A distância de Pearson baseia-se no desvio padrão da variável j , s_j , e é calculada de acordo com a Equação 1.13:

$$d_{ij} = \sqrt{\frac{\sum_{k=1}^K (x_{ik} - x_{jk})^2}{s_j^2}} \quad \text{Equação 1.13}$$

Depois de determinadas as distâncias, é necessário decidir sobre o número de grupos. Existem vários métodos, mas o algoritmo mais usado nos

espectros de FT-IR é o de Ward. Neste algoritmo, os grupos são agregados de modo a que exista um aumento mínimo do erro da soma dos quadrados, dentro do grupo definido. O erro da soma dos quadrados é definido como a soma dos desvios quadráticos de cada amostra do centro do seu grupo. A fórmula de cálculo do algoritmo de Ward para a distância entre uma nova amostra ou grupo k depende da média entre as distâncias individuais dos grupos A e B da amostra i , do seguinte modo:

$$d_{ki} = \frac{n_A + n_i}{n + n_i} d_{Ai} + \frac{n_B + n_i}{n + n_i} d_{Bi} - \frac{n_i}{n + n_i} d_{AB} \quad \text{Equação I.14}$$

Na equação I.14, d representa a distância entre duas amostras: A e/ou B e/ou i , n_A e n_B são o número de amostras no grupo A e B , respectivamente, e n é dado pela soma: $n = n_A + n_B$ ⁽¹¹⁸⁾.

No agrupamento hierárquico parte-se de n grupos de apenas um indivíduo cada, que vão sendo agrupados sucessivamente até se encontrar apenas um grupo que incluirá a totalidade dos n indivíduos⁽¹²⁸⁾.

A maioria dos artigos FT-IR referidos nesta dissertação usa esta ferramenta^(5-7, 13-15, 23, 24, 38, 50, 56, 65, 71, 74, 92, 93, 111, 113, 114, 124, 125, 136, 139, 141, 142, 151, 159, 160, 166, 169).

4.3. Modelos Supervisionados

4.3.1. Mínimos quadrados parciais (PLS)

O modelo de mínimos quadrados parciais (PLS) é também uma ferramenta muito utilizada em quimiometria⁽¹¹⁶⁾ e foi pela primeira vez descrita por Herman Wold em 1966. Tornou-se mais popular após o trabalho de Geladi e Kowalski em 1986⁽⁴⁶⁾. Em espectroscopia, o algoritmo PLS é baseado em

dois conjuntos de dados: os espectros (X) e a correspondente propriedade química ou física (Y). O objectivo da regressão PLS é estabelecer um modelo que permita a análise de amostras desconhecidas (determinação de uma propriedade física ou química a partir do espectro)⁽¹⁰²⁾ Olhando o modelo de uma forma simplificada, pode entender-se como uma regressão entre os componentes principais de ambos os blocos. No algoritmo PLS a extracção dos componentes principais, designados por variáveis latentes, é feita de uma forma distinta à do PCA.

O critério subjacente ao modelo PLS é o da maximização da correlação entre as variáveis independentes da matriz X e a variável dependente ou variáveis dependentes da matriz Y . O modelo calcula um conjunto de variáveis de modo a que a co-variância entre a matriz X e a matriz Y seja maximizada. Ambas as matrizes (X e Y) são decompostas em matrizes mais pequenas de acordo com a Equação I.15 e Equação I.16.

$$X = T \times P^T + E \quad \text{Equação I.15}$$

The diagram illustrates the decomposition of matrix X into components T , P^T , and E . Matrix X is shown as a rectangle with dimensions n (vertical) and p (horizontal). It is equal to matrix T (dimensions n by d) multiplied by matrix P^T (dimensions d by p), plus matrix E (dimensions n by p).

$$Y = U \times Q^T + F \quad \text{Equação I.16}$$

The diagram illustrates the decomposition of matrix Y into components U , Q^T , and F . Matrix Y is shown as a rectangle with dimensions n (vertical) and m (horizontal). It is equal to matrix U (dimensions n by d) multiplied by matrix Q^T (dimensions d by m), plus matrix F (dimensions n by m).

Nas equações apresentadas, X é uma matriz de $n \times p$ variáveis independentes, Y é uma matriz de $n \times m$ variáveis dependentes, T e U são matrizes $n \times d$ (componentes principais) que contêm linhas ortogonais, P é uma matriz $p \times d$ onde estão representados os pesos da matriz X , E corresponde à matriz $n \times p$ do erro residual da matriz X , Q é uma matriz $m \times d$ onde estão representados os pesos da matriz Y , e F corresponde à matriz $n \times m$ do erro residual da matriz Y ⁽¹¹⁸⁾.

No PLS a matriz de pesos é gerada através de um processo iterativo. A matriz W corresponde a uma matriz $d \times p$ dos pesos do PLS e está relacionada com a matriz de coeficientes do modelo geral de PLS, B , pela Equação I.17⁽¹¹⁸⁾:

$$B = W(P^T W)^{-1} Q^T \quad \text{Equação I.17}$$

O cálculo dos componentes, designados por variáveis latentes, é feito segundo o algoritmo de NIPALS⁽⁹⁶⁾.

Existem alguns artigos onde foi usado este algoritmo para o tratamento de resultados de FT-IR^(116, 122, 137).

4.3.2. Modelação independente de analogia de classes (SIMCA)

A análise de componentes principais é uma ferramenta que apenas permite fazer a análise exploratória dos dados, não permitindo a elaboração de modelos que sejam usados posteriormente para a identificação de uma nova amostra. O modelo SIMCA foi descrito em 1976⁽¹⁷¹⁾ e consiste na combinação de análises de PCA para cada classe que se pretende avaliar. Assim, para a construção deste modelo é preciso definir os espectros, a característica de

cada espectro que se pretende estudar (serótipo, identificação de estirpe, entre outras), designada neste trabalho por classe, e o número de componentes principais adequado.

Para cada classe q , é construído um modelo independente que identifique a amostra x .

$$x_{ij}^q = \bar{x}_j^q + \sum_{a=1}^{A_q} t_{ia}^q l_{ja}^q + e_{ij}^q \quad \text{Equação I.18}$$

Na Equação I.18, \bar{x}_j^q é a média da variável j na classe q , A_q é o número de componentes principais significativos na classe q , t_{ia}^q é o componente principal da amostra i no componente a e na classe q , p_{ja}^q é o peso da variável j no componente principal a na classe q e, e_{ij}^q é o erro residual da amostra i e da variável j ⁽¹¹⁸⁾.

Este método foi usado para avaliar os dados experimentais de discriminação de estirpes de cianobactérias⁽⁷⁴⁾ e de *Pseudomonas aeruginosa* e *Escherichia coli* em águas⁽²⁾.

4.3.3. Redes Neurais Artificiais (ANN)

Podem definir-se redes neuronais artificiais (ANN) como estruturas constituídas por elementos processados de uma forma simples, que estão interligados, (chamados neurónios artificiais ou simplesmente nodos) e que são capazes de ser processados de modo a obter um modelo que explique a realidade⁽¹²⁾.

As ANN são especialmente úteis quando se pretende modelar uma relação que se sabe à partida não ser linear⁽¹²⁾.

O nome desta ferramenta advém do facto de mimetizar a acção dos neurónios no cérebro. Um neurónio pode ser dividido em três partes: dendrites, corpo celular e axónio. A passagem de um sinal de um axónio para outro axónio é feita através de um espaço muito pequeno, por um processo que se chama de “sinapse”. As redes neuronais têm equivalentes ao neurónio: as dendrites e os axónios são nodos e as sinapses são os factores de conexão⁽¹²⁾. Nos neurónios, a transferência do sinal da extremidade de um axónio pela sinapse é simulada, na rede neuronal, pela multiplicação do sinal de entrada, x , por um peso sináptico, w , de modo a dar a resposta, y ⁽¹¹⁸⁾. Na Figura I.9 pode ver-se esquematizada a operação que ocorre num nodo. O nodo recebe os dados de entrada (dados que se pretendem analisar ou dados provenientes de outro nodo), x_i , combina-os através dos pesos das sinapses, w , e emite o resultado, após uma transformação adequada, na forma de sinal de saída, y_i .

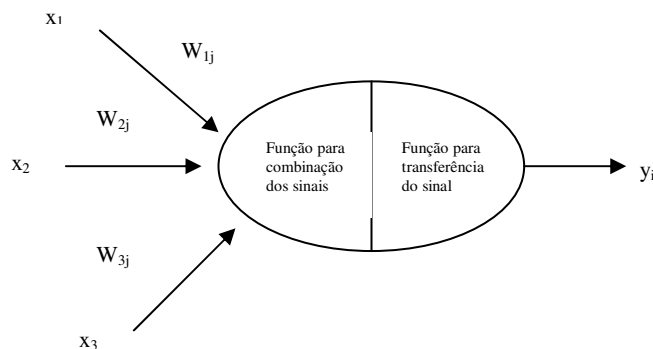


Figura I.9 – Esquema de um nodo de uma rede neuronal

Estas redes podem ter vários níveis interligados de neurónios artificiais que incluem um nível de valores de entrada e um de valores de saída. Os dados obtidos experimentalmente são introduzidos no nível de entrada e

são processados através de níveis intermédios (um ou mais), de modo a produzir um ou mais resultados.

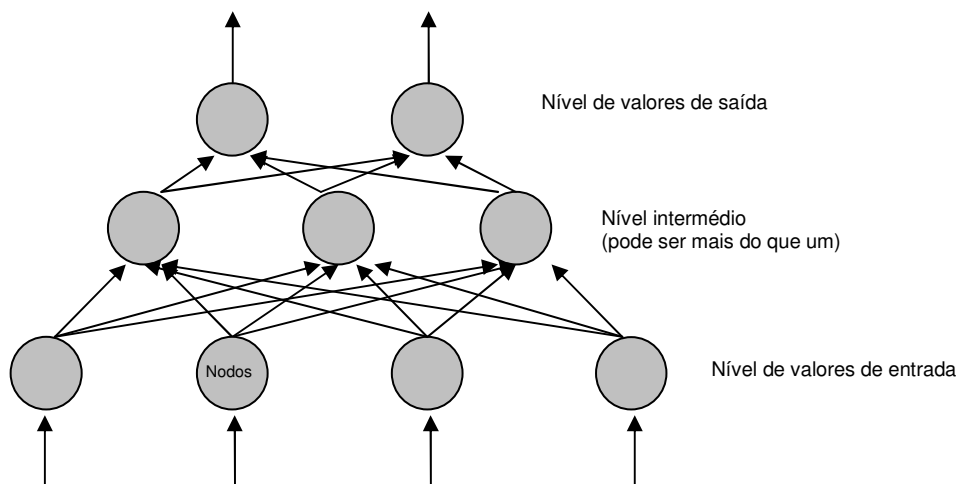


Figura I.10 – Esquema do exemplo de uma estrutura de uma rede neuronal

A discrepância entre os valores reais e os previstos pela rede permite o ajuste dos parâmetros da rede neuronal. Estes dois passos, previsão e ajuste, são repetidos até que se atinja o nível de precisão exigido. Depois de construído o modelo, este deve ser testado com outras amostras⁽¹⁰²⁾.

A primeira descrição do uso das redes neuronais para discriminação de espécies bacterianas é devida a Goodacre e seus colaboradores, neste caso usada para distinguir entre espécies clínicas de *Streptococcus* e *Enterococcus*⁽⁴⁹⁾. Seguiram-se outros estudos em que apenas com a aplicação de redes neuronais foi possível a distinção de determinadas características de estirpes, tais como a distinção entre estirpes de *S. aureus* sensíveis e resistentes à meticilina⁽⁵¹⁾, a identificação de espécies que provocam infecção urinária⁽⁵⁰⁾, a distinção entre *C. albicans* e *C. dubliniensis*⁽¹⁶¹⁾, a quantificação de bactérias patogénicas contaminantes de alimentos⁽⁵⁴⁾, a discriminação de 145 estirpes bacterianas ao nível do género e 24 leveduras ao nível da

espécie⁽¹⁶³⁾, e a identificação ao nível da espécie de *Pseudomonas* e estudo de susceptibilidade a anti-fúngicos de *C. albicans*⁽¹⁴⁰⁾.

4.4. Validação dos modelos

4.5.1. Validação cruzada

A validação cruzada é uma metodologia utilizada para a escolha do número de componentes principais ou variáveis latentes associada a um menor erro de previsão de um dado modelo.

Os modelos SIMCA e PLS deverão ser construídos de modo a que a identificação de uma amostra desconhecida seja a melhor possível. Quando existe um número de ensaios muito elevado, opta-se por dividir esse conjunto em dois subconjuntos: o primeiro (tipicamente 70%) usa-se para construir o modelo e o segundo (restantes 30%) usa-se para testar o modelo. Quando o número de ensaios é limitado, opta-se por uma estratégia de validação interna (validação cruzada). A operação de validação cruzada envolve um ou mais procedimentos internos de validação. Em cada procedimento faz-se a remoção de parte dos dados de calibração, utilizando-se os restantes para a construção do modelo (calibração) e consequente aplicação dos dados removidos (validação). Como esta operação não utiliza os mesmos dados para a construção e teste do modelo, permite estimar os modelos de uma forma mais próxima da óptima⁽¹⁰⁵⁾.

A raiz do erro quadrado médio da calibração (RMSEC) é calculada de forma semelhante ao RMSECV, com a diferença de que apenas um modelo é gerado com as N amostras, e testado com essas N amostras. Neste caso, o RMSEC decresce com o aumento do número de variáveis latentes, num

modelo de PLS. Isto acontece, uma vez que o melhor modelo corresponde àquele que contém o número máximo de variáveis latentes (modelo equivalente ao de regressão linear múltipla). Quanto maior a variância incorporada no modelo, melhor será a previsão correspondente. No entanto, este aumento do ajuste é aparente e só acontece especificamente para as amostras utilizadas. Em geral, o erro de generalização não acompanha geralmente o erro de calibração. A partir de um determinado número de variáveis latentes, o RMSECV deixa de ser aproximadamente proporcional ao RMSEC. O RMSEC não pode, por isso, ser usado para otimizar o número de variáveis latentes em modelos PLS⁽¹⁰¹⁾.

4.5. Análise dos modelos

4.5.1. Estatística de Hotelling (T^2)

A estatística de Hotelling (T^2) é utilizada para monitorizar os componentes do modelo PCA ou PLS. A Equação I.19 é usada para obter a estatística para uma amostra i correspondente a um componente principal ou variável latente t_i para o PCA ou PLS, respectivamente, para as primeiras A variáveis latentes.

$$T_A^2 = \sum_{i=1}^A \frac{t_i^2}{s_{t_i}^2} \quad \text{Equação I.19}$$

Nesta equação, s_{t_i} é a variância estimada da variável latente correspondente⁽⁸⁴⁾.

4.5.2. Estatística dos resíduos (estatística Q)

A estatística dos resíduos baseia-se no cálculo da soma dos quadrados dos resíduos para todas as variáveis. Assim, esta estatística mostra a falta de ajuste do modelo para cada amostra. A soma dos quadrados dos resíduos da amostra i pode ser calculada pela Equação 1.20.

$$Q_i = \sum_{j=1}^J \sum_{k=1}^K e_{ijk}^2 \quad \text{Equação 1.20}$$

Na Equação 1.20, e_{ijk} é um elemento da matriz $E_i(j \times k)$, que representa a matriz dos resíduos da amostra i . Os valores de Q_i podem-se aproximar de uma distribuição de χ^2 se os resíduos forem independentes e normalmente distribuídos:

$$Q_i \sim g\chi_h^2$$

Nesta equação, h são os graus de liberdade e g é um factor multiplicativo⁽⁹⁷⁾.

Os parâmetros g e h são calculados com base na média (m) e variância (ν) dos valores do erro do quadrado de previsão dos ensaios do conjunto nominal $\left(g = \frac{\nu}{2m}, h = \frac{2m^2}{\nu}\right)$ ⁽⁹⁶⁾.

Estas ferramentas podem ser usadas para analisar a existência de medidas atípicas (valores aberrantes ou “outliers”).

II. MÉTODOS

A parte prática desta dissertação foi realizada no Departamento de Microbiologia da Direcção de Comprovação da Qualidade do INFARMED, I.P.

1. Objecto de Estudo

As estirpes de *S. pneumoniae* incluídas no estudo pertencem ao Laboratório de Microbiologia da Faculdade de Medicina de Lisboa. Estas estirpes têm uma origem clínica e já foram avaliadas quanto à resistência ou sensibilidade à penicilina, pelo método CLSI, assim como também já foi determinado o serótipo pelo “método de xadrez”. Neste método, é determinado o grupo de serótipos ao qual a estirpe pertence e, depois com o auxílio dos soros dos serótipos presentes nesse grupo, é detectado o serótipo da estirpe a estudar⁽¹⁵⁵⁾.

1.1. Construção do banco de estirpes e criopreservação

As estirpes foram recebidas no Laboratório de Microbiologia do INFARMED, I.P. em placas de gelose de sangue. De cada placa foram retiradas todas as colónias que foram colocadas num criotubo esterilizado de 2 ml, contendo meio líquido de hidrolisado de caseína e de soja, e glicerol a 10%. Cada suspensão foi agitada e deixada à temperatura ambiente durante 30 minutos. Depois deste tempo os tubos foram colocados em caixas (Mr. Frosty[®]) contendo isopropanol e foram colocados num congelador a -70°C. Os contentores Mr. Frosty[®] permitem que a temperatura diminua de uma forma gradual, aumentando a taxa de recuperação das células aquando da repicagem para placas.

A descrição das 61 estirpes de *S. pneumoniae* estudadas encontra-se na Tabela II.1.

Tabela II.1 – Listagem das estirpes de *S. pneumoniae* incluídas no estudo e suas características de serótipo e susceptibilidade à penicilina.

	Serótipo					Total estirpes
	9V	14	19A	23F	NT	
Resistente	1999V0772	2001V0462	1999V1010	2000V1217	2002V1463	31
	2000V1226	2001V0761	2000V0324	2001V1072	2003V0416	
	2001V1277	2001V1096	2001V0406	2002V0067	2003V1113	
	2003V0262	2001V1152	2003V0585	2003V1481	(3)	
	2003V1528	2002V0929	2003V1485	2003V0063		
	(5)*	2003V0803	2003V1486	2003V1484		
		2003V0813	2003V1490	(6)		
		2003V1104	(7)			
		2003V1199				
		2003V1200				
		(10)				
Sensível	1999V0993	1999V0983	1999V0490	1999V0975	2000V0906	30
	2001V0180	2000V1210	2000V0193	2001V0919	2003V0319	
	2001V0182	2002V0488	2001V1222	2001V1123	2003V0583	
	2003V0785	2002V1462	2002V0701	2003V0088	2003V0787	
	(4)	2003V0790	2003V0033	2003V0148	2003V1203	
		2003V1494	2003V0076	(5)	(5)	
		(6)	2003V0577			
			2003V1112			
			2003V1489			
			2003V1495			
			(10)			
Total estirpes	9	16	17	11	8	61

*O número total de estirpes por cada serótipo encontra-se entre parêntesis. Com a designação NT encontram-se as estirpes cujo serótipo não foi determinado pelo “método de xadrez”.

1.2. Preparação das estirpes para análise

De cada criotubo a -70°C foi retirada uma pequena parte do gelo com uma ansa descartável e esterilizada. As estirpes foram depois semeadas num

meio de gelose de sangue (Biomérieux, ref^a 43001) e as placas foram incubadas a 35°C numa estufa com 5% de CO₂ por 24 ± 2h. Após o tempo de incubação, foi feita uma repicagem para uma segunda placa de gelose de sangue que foi incubada nas mesmas condições que a primeira. Todo o trabalho experimental foi feito a partir desta segunda placa.

O tratamento das estirpes foi feito de modo a ser possível a sua padronização, ou seja, o meio de crescimento teve sempre a mesma composição e as mesmas características.

2. FT-IR

Análise das amostras por absorvância no suporte de Selenato de zinco

2.1.1. Tratamento das estirpes

A partir das placas de gelose de sangue, preparadas de acordo com o descrito em 1.2, foram preparadas suspensões em água, de concentração aproximada de 3,5x10⁹ unidades formadoras de colónias por mililitro. Este procedimento foi realizado para cada estirpe, em três dias diferentes com lotes de meio de crescimento diferentes, de modo a avaliar a reprodutibilidade.

2.1.2. Análise das Suspensões

Cada suspensão foi colocada num suporte de Selenato de Zinco, em triplicado (de modo a avaliar a repetibilidade). Este suporte permite a colocação de 15 suspensões, tem uma posição adicional na qual não é colocada nenhuma amostra (posição “0”) que é usada para as leituras do ruído de fundo. Na Figura II.1, encontra-se um esquema do suporte de Selenato de Zinco usado.

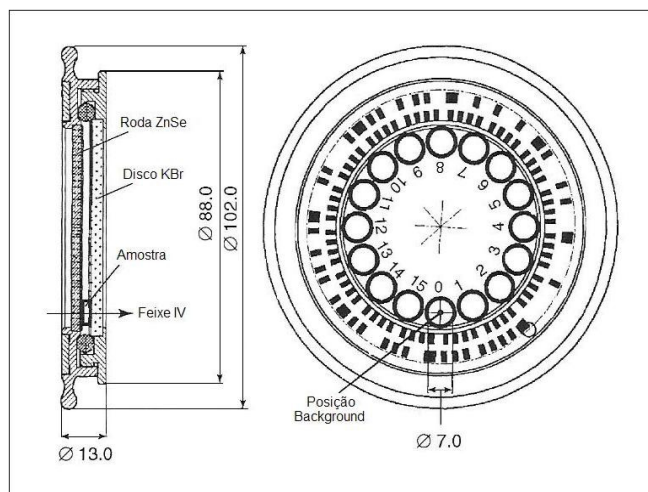


Figura II.1 – Esquema do suporte de Selenato de Zinco onde são colocadas as suspensões contendo os microrganismos a analisar. Este suporte é colocado no módulo IFS28/B do espectrofotômetro de FT-IR (adaptado de ⁽²⁷⁾).

Depois de aplicadas as suspensões, o suporte foi colocado numa estufa a 55°C durante 45 minutos de modo a que a suspensão ficasse imobilizada para que não restasse qualquer vestígio de água. Depois de secas as suspensões, o suporte foi fechado com uma tampa de cloreto de sódio e o suporte foi colocado no aparelho de FT-IR Equinox 55 acoplado ao Módulo B (Bruker Optik GmbH, Ettlingen, Alemanha) com uma fonte de infra-vermelho MIR e um detector de DTGS. O FT-IR encontrava-se anexo a um computador Pentium III com 256MB no qual estava instalado o software OPUS NT, versão 3.1.19, que faz a comunicação com o FT-IR. Os espectros foram adquiridos entre os 4000 e os 500 cm^{-1} com uma resolução de 6 cm^{-1} . Por cada espectro obtido, foram adquiridos 64 espectros que foram submetidos a uma média instrumental de modo a melhorar a razão sinal/ruído. Antes da leitura de cada amostra, foi feita uma leitura do branco cujo espectro foi automaticamente subtraído ao espectro da amostra. O tempo de aquisição de cada espectro foi de aproximadamente 2 minutos e os espectros são apresentados na forma de absorvância.

2.1.3. Avaliação da importância da escolha do quadrante de crescimento na análise de FT-IR

De modo a avaliar a importância de se usarem apenas colônias pertencentes ao terceiro quadrante, ou a qualquer um dos outros, foi feito um estudo com uma estirpe (2003V1490\$su19A) da qual apenas se tomaram colônias pertencentes ao primeiro quadrante para uma suspensão, ao segundo quadrante para outra suspensão e ao terceiro quadrante para uma terceira suspensão. Estas três suspensões foram preparadas em duplicado e cada uma destas foi colocada em triplicado numa roda de selenato de zinco. A aquisição dos espectros foi feita de acordo com o descrito no ponto anterior.

2.2. **Análise das amostras por ATR**

A partir de uma placa de gelose de sangue da estirpe 2003V1490\$su19A, foram retiradas várias colônias que foram colocadas no acessório ATR. Os espectros foram adquiridos entre os 4000 e os 400 cm^{-1} , com resolução de 4 cm^{-1} . Cada espectro obtido é a média de 32 espectros de modo a melhorar a razão sinal/ruído. Antes da leitura da amostra, foi feita uma leitura do ruído de fundo cujo espectro foi automaticamente subtraído ao espectro da amostra.

3. **Cromatografia Gasosa**

A análise dos ácidos gordos foi realizada segundo a metodologia definida pela empresa MIDI que desenvolveu o método⁽¹⁾, e que comercializa uma base de dados para identificação de microrganismos com o nome

Sherlock[®]. Neste caso não foi importante a identificação correcta das estirpes analisadas mas sim, a identificação dos ácidos gordos constituintes de cada estirpe. De modo a rastrear os problemas inerentes ao cromatógrafo, à coluna ou ao tubo protector⁽¹¹²⁾, antes de analisar qualquer amostra, foi injectado um padrão específico para o efeito. Este padrão é composto por uma mistura de vários ácidos gordos saturados de 9 a 20 átomos de carbono, e de 5 hidroxácidos, que originaram sinais que foram comparados com os que o programa tinha definido. Consoante o desvio, os parâmetros inseridos na base de dados foram alterados de modo a poder dar a resposta pretendida para o padrão, seguindo depois o mesmo procedimento para as amostras. Este padrão foi analisado entre cada 10 amostras.

De modo a controlar a eficiência da extracção, em cada análise foi incluída nas amostras a analisar, um microrganismo com uma composição em ácidos gordos complexa e conhecida, a *Pseudomonas aeruginosa* ATCC 9027.

3.1. Preparação das estirpes

Os pneumococos a estudar foram tratados de acordo com o descrito nos pontos 1.1 e 1.2 deste capítulo. Por cada conjunto de estirpes a analisar foi também incorporada a estirpe *Pseudomonas aeruginosa* ATCC 9027. A partir da segunda placa de gelose de sangue foram retiradas, para um tubo, aproximadamente 10 mg de colónias, com o cuidado de retirar apenas do terceiro quadrante de crescimento, ou aquele imediatamente anterior ao correspondente às colónias isoladas.

3.2. Saponificação

O objectivo da saponificação é lisar as células, havendo a libertação dos ácidos gordos que são convertidos nos seus sais de sódio.

Pipetou-se $1,0 \pm 0,1$ ml do reagente GC #1 (composição descrita no Anexo I) para cada tubo de ensaio com as estirpes a analisar e para um tubo que não tinha colónias e que serviu de branco do ensaio. Agitou-se no vortex durante 5 a 10 segundos e colocaram-se os tubos num banho com temperatura entre 95 e 100°C durante 5 minutos. Após este tempo foi feita uma agitação no vortex e os tubos foram novamente colocados no banho durante 25 minutos. Depois da incubação, os tubos foram colocados num recipiente com água fria.

3.3. Metilação

O objectivo da metilação é a conversão dos ácidos gordos, que se encontram na forma de sais, a metilésteres, o que permite o aumento da volatilidade.

Depois de arrefecidos, adicionaram-se aos tubos $2,0 \pm 0,2$ ml do reagente GC #2 (descrição da composição no Anexo I). Homogeneizou-se no vortex e colocou-se num banho a $80 \pm 1^\circ\text{C}$ durante 10 ± 1 minutos. Arrefeceram-se os tubos num banho de água fria.

3.4. Extracção

O objectivo deste passo é extrair os metilésteres de ácidos gordos da fase aquosa para a fase orgânica, uma vez que fases aquosas danificam as colunas de GC.

Adicionaram-se aos tubos $1,25 \pm 0,1$ ml do reagente GC #3 (descrição da composição no Anexo I) e colocaram-se num agitador orbital durante 10 minutos. Findo este tempo e, com o auxílio de uma pipeta Pasteur de vidro, foi eliminada a fase aquosa.

3.5. Purificação

O objectivo da fase da purificação é remover os ácidos gordos livres e os reagentes residuais do extracto orgânico, ficando apenas na fase orgânica os metilésteres de ácidos gordos.

Foram colocados $3,0 \pm 0,3$ ml do reagente GC #4 (descrição da composição no Anexo I) em cada tubo e os tubos foram colocados num agitador orbital durante 5 minutos. Depois, foram adicionadas algumas gotas do reagente GC #5 (descrição da composição no Anexo I) de modo a desfazer a emulsão. Com uma pipeta de Pasteur de vidro, foram transferidos 2/3 da fase orgânica para um vial de GC. Selou-se o vial.

3.6. Análise por GC

As condições cromatográficas foram as seguintes:

Injecção automática de 2 μ L de amostra num injector que se encontrava a 250 °C. A amostra foi introduzida na coluna numa proporção 1:50. O gás de arrasto usado foi o hidrogénio e o gás “make up” o azoto. A coluna usada foi uma coluna ultra-capilar do fornecedor Agilent com as dimensões 25m x 0,200 mm com uma espessura de enchimento de (5%-fenil)-metilpolisiloxano de 0,33 μ m. Foram programadas duas rampas de temperatura: a primeira que se iniciou a 170°C e terminou a 288°C com uma subida de 28°C/minuto e a segunda que terminou aos 310°C à qual ficou

durante 1,25 minutos, com um aumento de temperatura de 60°C/minuto. Os resultados foram recolhidos num detector FID para o qual é necessário ar sintético puro. Os sinais obtidos foram analisados pelo software “Chemstation” e foram processados pelo software “Sherlock®” que apresentou um relatório com as informações relevantes.

4. Análise de Resultados

4.1. Teste de Qualidade para os espectros de FT-IR

Os espectros foram avaliados quanto à sua qualidade no que diz respeito à relação sinal/ruído, à intensidade do vapor de água e à diferença entre o máximo e o mínimo de absorvância. Os critérios usados para avaliar estes parâmetros foram definidos pelo fabricante do equipamento de FT-IR (Bruker) e encontram-se descritos na Tabela II.2:

Tabela II.2 – Limites para os parâmetros avaliados no teste de qualidade dos espectros

	Amplitude x1	Amplitude x2	Amplitude x3	Amplitude x4	Amplitude x5	Amplitude x6
X inicial	2100	1700	1200	2100	1847	2300
X final	1600	1600	960	2000	1837	2000
Máximo delta y	0,2					
Mínimo delta y	2					
Ruído				0,00015		
Sinal/Ruído		150	40			
Água máx.					0,0003	
Sinal/água		100	20			
Interferências						0,0002

Uma explicação pormenorizada do significado dos parâmetros desta tabela encontra-se no Anexo II.

4.2. Avaliação dos resultados de GC

Os resultados aparecem sob forma de cromatograma e numa tabela onde apenas se encontram descritos os picos detectados no cromatograma. Por cada pico encontra-se o tempo de retenção, a resposta, a relação área/altura do pico, a identificação do pico e a percentagem da resposta relativamente ao conjunto de picos encontrados. Um parâmetro importante no relatório é a resposta total, esta deverá ser superior a 50000 para se poder aceitar o resultado. No caso de resultados com resposta inferior a este valor, o frasco contendo a amostra foi colocado sob corrente de azoto de modo a evaporar parte do solvente, concentrando os ácidos gordos e, consequentemente, aumentando a resposta. No relatório da análise também se encontra a identificação da estirpe analisada, por comparação à base de dados do Sherlock[®]. Associada a cada identificação listada encontra-se o índice de semelhança (SI). Esta identificação não foi usada neste trabalho, uma vez que todas as estirpes analisadas foram identificadas como *S. pneumoniae* no Laboratório de Microbiologia da Faculdade de Medicina de Lisboa e que, para uma correcta identificação pelo Sherlock[®] a metodologia de crescimento das estirpes deveria ser exactamente igual à usada na elaboração da base de dados, o que não aconteceu: as estirpes foram incubadas numa estufa a 35°C com 5% de CO₂ em vez de serem incubadas a 35°C numa estufa sem CO₂. Assim, apenas se teve em conta a tabela com a identificação dos ácidos gordos para cada amostra.

4.3. Estatística

O tratamento dos resultados foi realizado com o auxílio do software Matlab, versão 6.5 (The Mathworks, Inc., Natick, Massachussets) com a “PLS Toolbox”⁽¹⁷⁰⁾, que corre em ambiente “Microsoft Windows XP”.

4.4. Cálculo da Repetibilidade e da Reprodutibilidade

O cálculo da repetibilidade baseia-se em resultados obtidos num mesmo dia resultantes de diferentes medidas de uma mesma suspensão. Este parâmetro tem em conta apenas as variações do equipamento de medida. A reprodutibilidade calcula-se com os resultados obtidos em dias diferentes para uma mesma estirpe, nas mesmas condições, e cuja suspensão foi preparada do mesmo modo para os vários dias.

A fórmula de cálculo é a mesma para a repetibilidade e para a reprodutibilidade, sendo que o que varia são os dias aos quais correspondem os resultados: do mesmo dia para a repetibilidade, ou de dias diferentes para a reprodutibilidade. A repetibilidade e a reprodutibilidade são avaliadas pela determinação do coeficiente de correlação entre pares de espectros, $r_{y_1 y_2}$ ⁽¹⁴⁶⁾, (Equação II.1) e pelo cálculo de um índice de diferenciação, $D_{y_1 y_2}$ ⁽⁶⁶⁾ (Equação II.2).

$$r_{y_1 y_2} = \frac{\sum_{i=1}^n y_{1i} y_{2i} - n \overline{y_1} \overline{y_2}}{\sqrt{\sum_{i=1}^n y_{1i}^2 - n \overline{y_1}^2} \times \sqrt{\sum_{i=1}^n y_{2i}^2 - n \overline{y_2}^2}} \quad \text{Equação II.1}$$

$$D_{y_1 y_2} = (1 - r_{y_1 y_2}) \times 1000 \quad \text{Equação II.2}$$

Nas equações referidas y_{1i} e y_{2i} são os valores de absorvância dos dois espectros a comparar, n é o número de pontos da gama de números de onda e $\overline{y_1}$ e $\overline{y_2}$ são as médias aritméticas dos valores y_1 e y_2 .

O valor de $D_{y_1y_2}$ pode estar compreendido entre 0 e 2000. Se for 0 pode-se considerar que os espectros são idênticos, se for 1000 consideram-se que os espectros não estão correlacionados, e acima de 2000 estão negativamente correlacionados^(59, 60, 108). Na literatura estão descritos valores de 0,4-0,8 para a repetibilidade e 7-10 para a reprodutibilidade, para protocolos normalizados. Os limites para $D_{y_1y_2}$ foram obtidos após a análise da primeira derivada dos espectros de FT-IR, numa gama espectral de 900 a 1200 cm^{-1} . Se se pretenderem usar estes parâmetros terá de ser seguido o pré-processamento descrito^(59, 60).

III. RESULTADOS E DISCUSSÃO

1. Espectros

Os espectros de FT-IR são usualmente representados usando como eixo das abcissas o número de onda e como eixo das ordenadas a absorvância. Um exemplo de um espectro obtido é o seguinte:

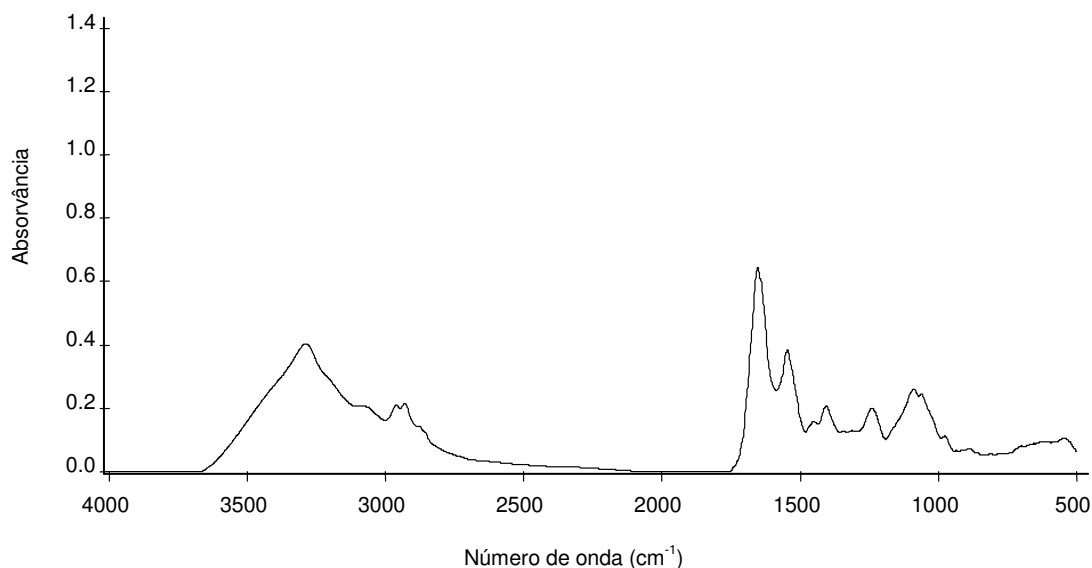


Figura III.1 – Espectro de FT-IR para uma estirpe de *S. pneumoniae*.

Como se pode observar na figura, o espectro de FT-IR de *S. pneumoniae* é bastante complexo e está de acordo com o representado na Figura I.6.

Um exemplo dos espectros obtidos para os 5 serótipos analisados neste trabalho está representado na Figura III.2.

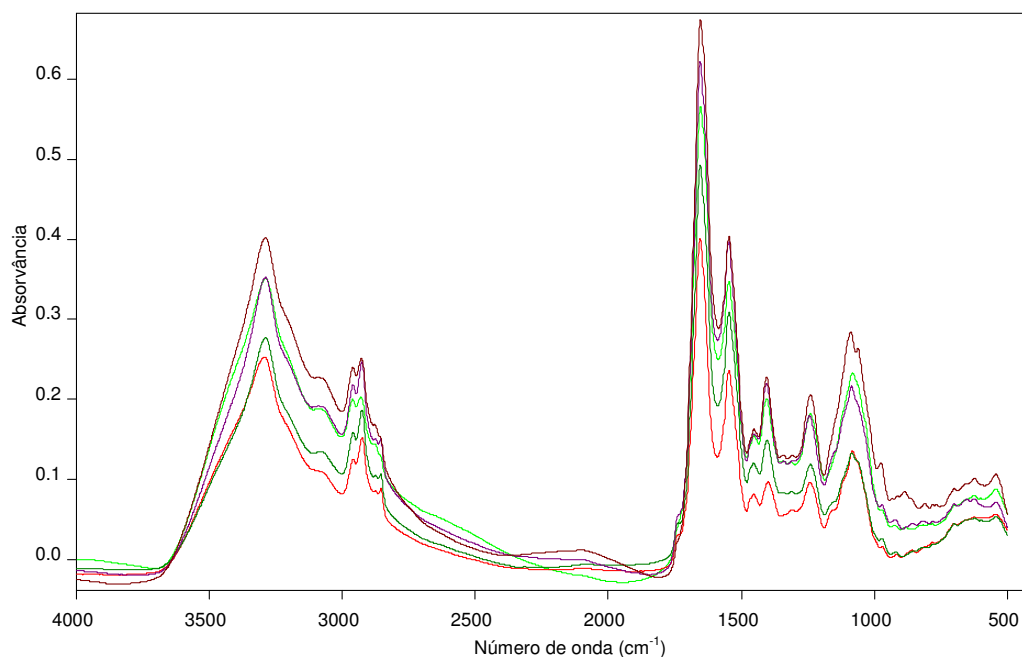


Figura III.2 – Espectro de FT-IR para 5 estirpes de *S. pneumoniae* de 5 serótipos distintos. A (—) encontra-se um espectro de uma estirpe de serótipo 9V, a (—) uma estirpe de serótipo 14, a (—) uma estirpe de serótipo 19A, a (—) uma estirpe de serótipo 23F e a (—) uma estirpe sem serótipo determinado (NT).

Observando a Figura III.2, pode-se constatar que não existe uma diferença visual entre os cinco serótipos. Por este facto, terá de ser usada análise multivariada para tentar explicar as diferenças entre os serótipos.

Na Figura III.3 encontra-se um exemplo de dois espectros de uma estirpe sensível e de uma estirpe resistente à penicilina.

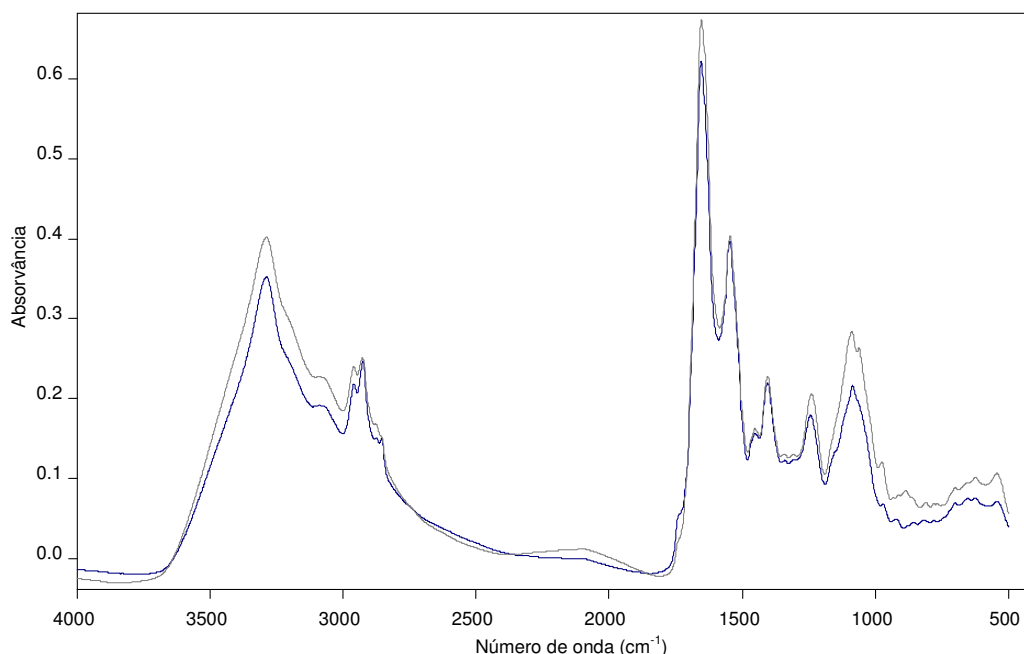


Figura III.3 – Espectros de FT-IR de estirpes de *S. pneumoniae* de uma estirpe sensível (—) (1999V0490) e outra estirpe resistente (—) à penicilina (1999V0772).

2. Teste de Qualidade

Para as estirpes analisadas obteve-se um total de 1380 espectros, dos quais apenas 542 satisfizeram o teste de qualidade. O teste de qualidade foi feito com auxílio do software OPUS NT, versão 3.1.19 e o resultado que se obteve foi uma tabela que contém os valores alvo para cada parâmetro (intervalo de absorvância, ruído, relação sinal / ruído, presença de água e relação sinal / água) e os valores obtidos para o espectro em questão. Dois exemplos de tabelas, um de um espectro que satisfaz e outro de um espectro que não satisfaz o teste de qualidade, encontram-se nas Figura III.4 e Figura III.5.

	min	max	measured
Absorption (X-Range 1):	0.200000	2.000000	0.685665
Noise (X-Range 4):	0.000000	0.000150	0.000119
Signal/Noise (X-Range 2):	150.000000	0.000000	215.322295
Signal/Noise (X-Range 3):	40.000000	0.000000	42.302606
Water Vapour (X-Range 5):	0.000000	0.000300	0.000057
Signal/Water (X-Range 2):	100.000000	0.000000	453.895366
Signal/Water (X-Range 3):	20.000000	0.000000	89.173101
Fringes (X-Range 6):	0.000000	0.000200	0.000052

Figura III.4 - Resultado do teste de qualidade de um espectro de FT-IR que cumpre os critérios nos parâmetros definidos

	min	max	measured
Absorption (X-Range 1):	0.200000	2.000000	0.280349
Noise (X-Range 4):	0.000000	0.000150	0.000062
Signal/Noise (X-Range 2):	150.000000	0.000000	158.013631
Signal/Noise (X-Range 3):	40.000000	0.000000	39.755037
Water Vapour (X-Range 5):	0.000000	0.000300	0.000032
Signal/Water (X-Range 2):	100.000000	0.000000	310.035155
Signal/Water (X-Range 3):	20.000000	0.000000	78.002505
Fringes (X-Range 6):	0.000000	0.000200	0.000030

Figura III.5 – Resultado do teste de qualidade de um espectro de FT-IR que não pode ser usado pois não cumpre os critérios definidos para o teste de absorvância e para o teste da razão sinal/água.

Graficamente, um exemplo de um espectro que cumpre os critérios definidos no teste de qualidade e de um espectro que não cumpre os referidos critérios, encontra-se na Figura III.6. Nesta figura foi colocado um rectângulo a sinalizar a zona do espectro responsável pela falha no teste de qualidade. A verde está o espectro que não passa o referido teste e que apresenta uma banda pequena numa zona onde, para os microrganismos, não existem descrições da existência de picos e que, por esta razão, serve como referência para a medida do ruído de fundo.

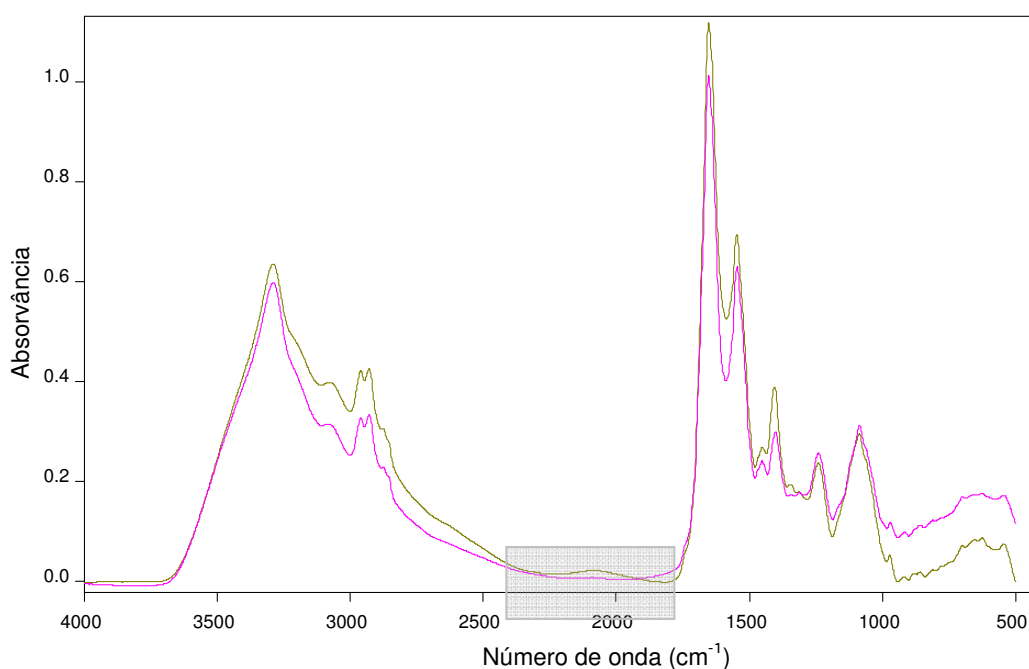


Figura III.6 – Exemplo de dois espectros da estirpe 2003V0076\$u19A onde o espectro a (—) passa o teste de qualidade e o espectro (—) não passa o teste.

Verificou-se que a taxa de insucesso no teste de qualidade foi bastante elevada no início do trabalho (12 análises em 15 espectros analisados não cumpriram os parâmetros do teste de qualidade), tendo sido melhorada no fim (6 espectros não cumpriram os parâmetros do teste de qualidade por 45 análises). No início, a ansa, depois de esterilizada à chama, era arrefecida na placa de gelose de sangue onde as estirpes cresceram. Uma vez que a taxa de sucesso foi tão reduzida (20%), passou-se a arrefecer a ansa em água esterilizada. A partir desta altura a quantidade de espectros que satisfizeram o teste de qualidade foi bastante mais elevada (87%).

De modo a avaliar a importância do teste de qualidade na identificação dos serótipos, fez-se uma análise de componentes principais do conjunto total de espectros (38) obtidos para a estirpe de *S. pneumoniae* 2003V0033\$u19A, conjunto este que engloba espectros que satisfazem e não satisfazem o teste de qualidade. Para a análise de componentes principais usou-se como pré-processamento o filtro de Savitzky-Golay (9,2,2), a função

SNV e a representação do componente principal número 3 em função do componente principal número 2. O resultado obtido encontra-se na Figura III.7.

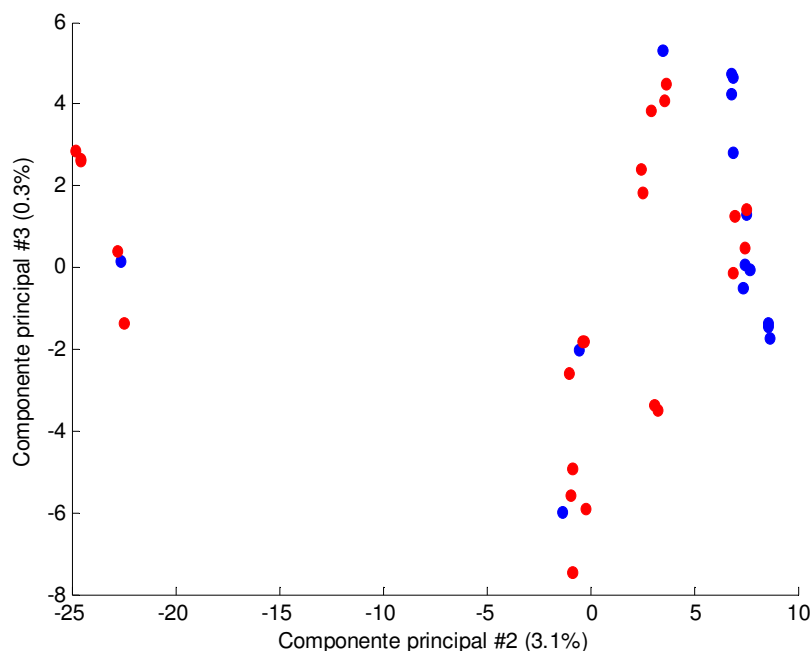


Figura III.7 – Análise de componentes principais para todos os espectros de *Streptococcus pneumoniae* que satisfizeram e não satisfizeram o teste de qualidade da estirpe 2003V0033\$su19A. Os pré-processamentos usados foram: SNV e filtro de Savitzky-Golay (9,2,2). Representado com o símbolo (●) estão representadas as estirpes que não passam o teste de qualidade e com o símbolo (●) as estirpes que passam o teste.

Como se pode observar, o teste de qualidade, nas condições de pré-processamento e zona espectral usada na análise dos dados das secções seguintes, não tem influência nos resultados uma vez que não existe uma separação evidente dos espectros que satisfazem e que não satisfazem o teste de qualidade. No entanto, na análise dos resultados não foram usados os espectros que não cumpriram os critérios definidos no teste de qualidade.

3. Cálculo da Repetibilidade e Reprodutibilidade dos Espectros

Foi calculada a repetibilidade e reprodutibilidade para várias estirpes. Aqui apresenta-se o exemplo para duas estirpes: 2001V0462\$su14 e

2003V1481\$su23F. De acordo com o descrito na secção "II. Métodos", o cálculo da repetibilidade e reprodutibilidade é feito de acordo com as Equações II.1 e II.2. O valor que se determina é um valor D para uma matriz em que se cruzam todos os espectros de uma mesma estirpe, ou seja, para as três réplicas de um dia e para os três dias distintos, e o que se obtém é uma matriz de 8 linhas por 9 colunas. Existem limites descritos para avaliar a repetibilidade e reprodutibilidade para a zona espectral da parede e para o pré-processamento de espectros com filtro de Savitzky-Golay (9,2,1). Estes limites são:

- para os cruzamentos da matriz de um mesmo dia o valor de D deverá estar entre 0,4 e 0,8
- para estirpes analisadas em dias diferentes o valor de D deverá ser superior a 7 e inferior a 10.

Para a estirpe 2001V0462\$su14, os valores de D encontram-se na Tabela III.1.

Tabela III.1– Tabela de valores de valor D para os vários cruzamentos de espectros por dia de análise para a estirpe 2001V0462\$su14.

	04/12/2006	05/12/2006	04/12/2006	05/12/2006	04/12/2006	05/12/2006	18/01/2007	18/01/2007	18/01/2007
04/12/2006	0	0,91	0,57	0,71	0,53	1,68	0,47	0,43	0,42
05/12/2006	0	0	0,97	0,11	1,14	0,57	0,92	0,81	0,72
04/12/2006	0	0	0	0,72	0,16	1,95	1,32	0,39	0,61
05/12/2006	0	0	0	0	0,86	0,64	0,92	0,68	0,70
04/12/2006	0	0	0	0	0	1,75	1,25	0,51	0,78
05/12/2006	0	0	0	0	0	0	1,50	1,72	1,66
18/01/2007	0	0	0	0	0	0	0	0,46	0,33
18/01/2007	0	0	0	0	0	0	0	0	0,10

Como se pode observar nos vários cruzamentos da Tabela III.1, não existem valores de D acima dos limites. Para esta estirpe não foram encontrados espectros aberrantes aquando da análise de PCA pela estatística de Hotelling e pela estatística dos resíduos (na secção 5.1.1).

Fazendo o mesmo tratamento para a estirpe 2003V1481\$su23F, o que se obtém é o descrito na Tabela III.2.

Tabela III.2 - Tabela de valores de valor D para os vários cruzamentos de espectros por dia de análise para a estirpe 2003V1481\$su23F.

	02/11/2006	03/11/2006	07/12/2006	02/11/2006	03/11/2006	07/12/2006	02/11/2006	03/11/2006	07/12/2006
02/11/2006	0	29,21*	42,08	0,34	29,71	41,73	1,99*	29,56	41,48
03/11/2006	0	0	5,78	30,16	0,05	5,89	27,51	0,11	6,53
07/12/2006	0	0	0	43,41	6,36	0,04	40,21	6,73	0,32
02/11/2006	0	0	0	0	30,61	43,10	1,13	30,59	43,12
03/11/2006	0	0	0	0	0	6,49	27,95	0,09	7,22
07/12/2006	0	0	0	0	0	0	39,94	6,82	0,27
02/11/2006	0	0	0	0	0	0	0	28,00	40,42
03/11/2006	0	0	0	0	0	0	0	0	7,39

*Os valores sublinhados a amarelo correspondem aos que foram eliminados por estarem no intervalo de confiança de outro serótipo e a verde está o espectro considerado aberrante

Como se pode observar na Tabela III.2, existem alguns valores que estão acima dos limites para o valor de D . Estes espectros são coincidentes com os que foram considerados aberrantes pela estatística de Hotelling e pela estatística dos resíduos aquando da análise de PCA (descrito na secção 5.1.1), e os que se eliminaram por estar no intervalo de confiança de outro serótipo que não o seu. De qualquer modo pode ver-se que a repetibilidade dos espectros do dia 02/11/2006 não é boa devido ao terceiro espectro desse dia mas, entre o primeiro e o segundo, o valor de D está dentro do limite. Existe uma boa reprodutibilidade entre os dias 03/11/2007 e 07/12/2007 e a repetibilidade para cada um dos dias também está conforme.

As tabelas referidas nesta secção são exemplos de duas determinações do valor de D , mas este valor foi determinado para outras 21 estirpes. Para alguns casos, os valores de D determinados reforçaram a eliminação dos espectros aberrantes pela estatística de Hotelling e pela estatística dos resíduos, tal como foi apresentado para a estirpe 2003V1481\$su23F. Para outras estirpes em que não foram eliminados

espectros aberrantes existem valores de D acima dos valores limite definidos. Como estes valores foram definidos em condições muito particulares e não para o *S. pneumoniae*, aceitaram-se espectros com uma pequena diferença entre o limite superior e o valor definido para D .

4. Verificação da Importância da Escolha dos Quadrantes

Para algumas das suspensões analisadas em FT-IR, foi necessária a recolha de colónias pertencentes aos três primeiros quadrantes, pois as colónias pertencentes ao terceiro quadrante não foram suficientes para a preparação das soluções a analisar. Assim, deve-se avaliar a influência que as diferentes fases de crescimento das colónias podem ter nos espectros de FT-IR, uma vez que, segundo a técnica descrita pela Bruker, apenas se devem usar colónias pertencentes ao terceiro quadrante. Para este estudo prepararam-se várias placas da estirpe 2003V1490\$su19A, da qual apenas se usaram colónias pertencente ao primeiro quadrante para uma suspensão, ao segundo quadrante para outra suspensão e ao terceiro quadrante para uma terceira suspensão. Estas três suspensões foram preparadas em duplicado e cada uma destas foi colocada em triplicado numa roda de selenato de zinco. O processamento dos resultados foi o mesmo que o descrito na avaliação dos serótipos, isto porque se pretende avaliar a importância dos quadrantes num determinado modelo. Não se considerou o tratamento dos dados aquando da avaliação da susceptibilidade à penicilina uma vez que não houve nenhum modelo aceitável.

Para a avaliação dos serótipos, o pré-processamento optimizado foi conseguido com o filtro de Savitzky-Golay (9,2,2) e SNV. A representação gráfica onde se evidenciam as diferenças dos serótipos é aquela em que se

projecta o componente principal 3 *versus* o componente principal 2, tal como descrito no ponto 5.1.1.

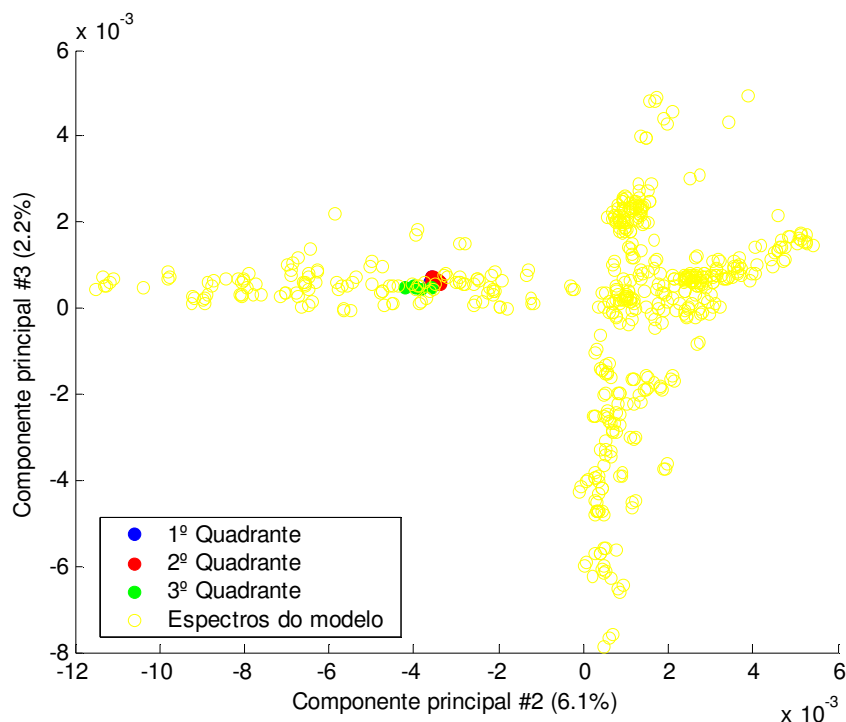


Figura III.8- Análise de PCA para 18 espectros da estirpe 2003V1490\$u19A de *S. pneumoniae*. Os espectros usados foram submetidos a um pré-processamento de filtro de Savitzky-Golay (9,2,2) e SNV. A zona do espectro usada correspondeu aos números de onda compreendidos entre 900 e 1185 cm^{-1} .

Como se pode observar na figura, não existe uma separação nítida dos espectros de acordo com o quadrante de crescimento. Assim, a recolha de colónias pertencentes a outros quadrantes, que não o terceiro, não deverá ter afectado os resultados.

5. Serotipagem

5.1. FT-IR

Foram testadas estirpes pertencentes aos 4 serótipos identificados (14, 9V, 19A e 23F) e estirpes cujo resultado da serotipagem pelo “método de

xadrez” não foi conclusivo. O resultado não conclusivo pode ocorrer devido a uma expressão insuficiente de polissacáridos da cápsula para reagir com o serótipo adequado ou, o serótipo não se encontrar representado no método usado.

Com os espectros de FT-IR obtidos pretende avaliar-se a possibilidade, ou não, da utilização da técnica de FT-IR para a serotipagem. Para isso, é necessária a elaboração de modelos estatísticos cuja identificação obtida pelo “método de xadrez” seja coincidente com a prevista pelo modelo. Com este fim, foram testadas várias abordagens quimiométricas como se apresenta em seguida.

5.1.1. PCA

O primeiro passo para a realização do PCA é a determinação do número de componentes adequados ao estudo. Este deverá ser o menor número possível mas com o máximo de informação. Uma das técnicas para determinar este número baseia-se na representação gráfica dos valores próprios em função do número de componentes principais.

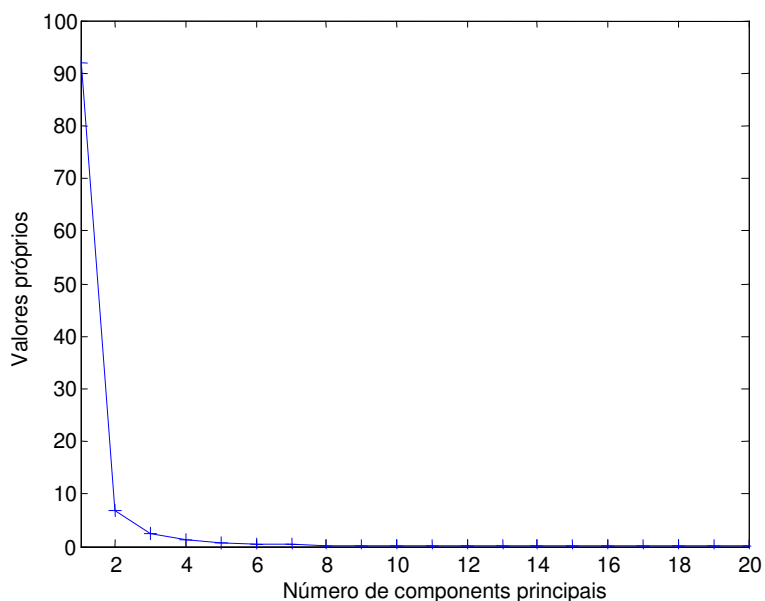


Figura III.9 – Representação gráfica dos valores próprios em função do número de componentes principais.

De acordo com a regra que indica que o número de componentes principais a usar será aquele a partir do qual se observa uma descida repentina no declive das rectas do gráfico, no caso da Figura III.9, dever-se-iam escolher apenas 2 componentes principais.

O método mais adequado para a determinação de componentes principais é aquele que faz uma análise discriminante do número de componentes e que dá como resultado um gráfico do erro associado ao número de componentes principais. Um exemplo deste tipo de gráficos encontra-se na Figura III.10.

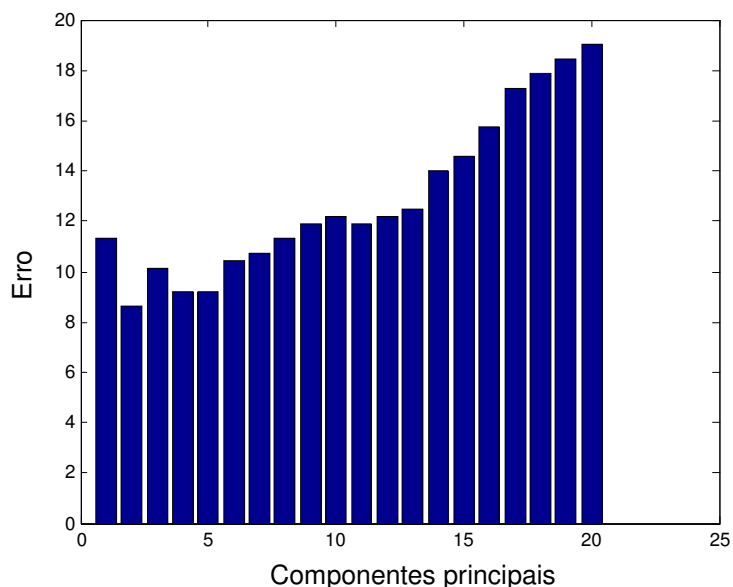


Figura III.10 – Gráfico de barras que mostra o erro associado ao número de componentes principais. A barra menor corresponde ao número de componentes principais que se deve escolher para a análise de PCA.

De acordo com a Figura III.10, o número de componentes principais associados ao menor erro é 2. Se se realizar um gráfico de componentes principais aos dados dos espectros de FT-IR tal como eles foram obtidos, usando 2 componentes principais, obtém-se o representado na Figura III.11.

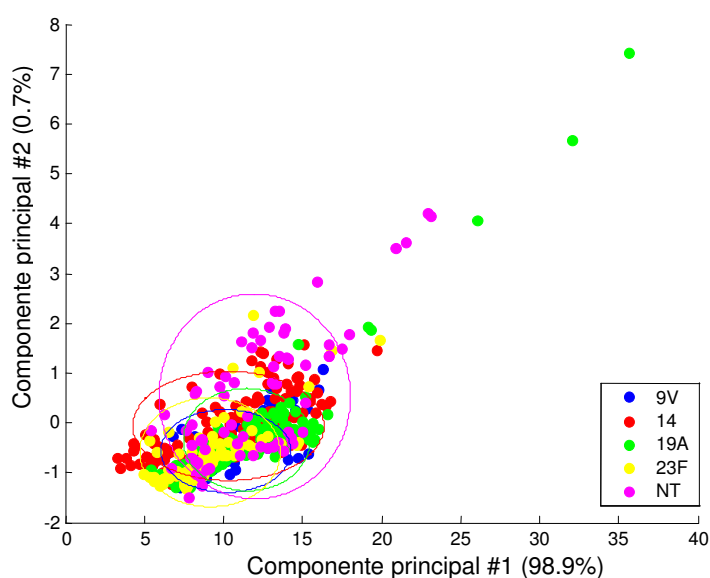


Figura III.11 – Análise de componentes principais para os 542 espectros, tal como foram obtidos pela análise de FT-IR, para estirpes de *S. pneumoniae*. Cada cor está relacionada com um dos 5 serótipos analisados.

Como se pode ver pela Figura III.11, uma análise de componentes principais aos espectros, tal como foram adquiridos, não é suficiente para distinguir as estirpes de acordo com o seu serótipo.

Uma vez que os espectros foram adquiridos em dias diferentes e de estirpes crescidas em meios de cultura de lotes diferentes, é natural que exista uma variabilidade de resultados. Assim, devem-se aplicar tratamentos a todos os espectros de modo a minimizar essas diferenças, o que se consegue com pré-processamentos. Foram testados vários pré-processamentos às amostras, de modo a verificar qual seria o mais adequado para se obter os resultados que se pretendiam, ou seja, uma representação que permita distinguir os 5 serótipos.

Representando a análise de componentes principais para todo o espectro e vários pré-processamentos (filtro de Savitzky-Golay (17,2,1) + remoção da média; filtro de Savitzky-Golay (17,2,2) + remoção da média; SNV + remoção da média, filtro de Savitzky-Golay (9,2,2) + SNV e MSN) para 2 componentes principais (gráfico não apresentado) e para 3 componentes principais, o melhor resultado obtém-se para a representação gráfica do componente principal 3 em função do componente principal 2, mas não se conseguindo ainda distinguir as estirpes pelo seu serótipo, como se pode ver na Figura III.12. Nesta figura cada ponto corresponde a um espectro e os círculos correspondem ao intervalo de confiança para cada serótipo.

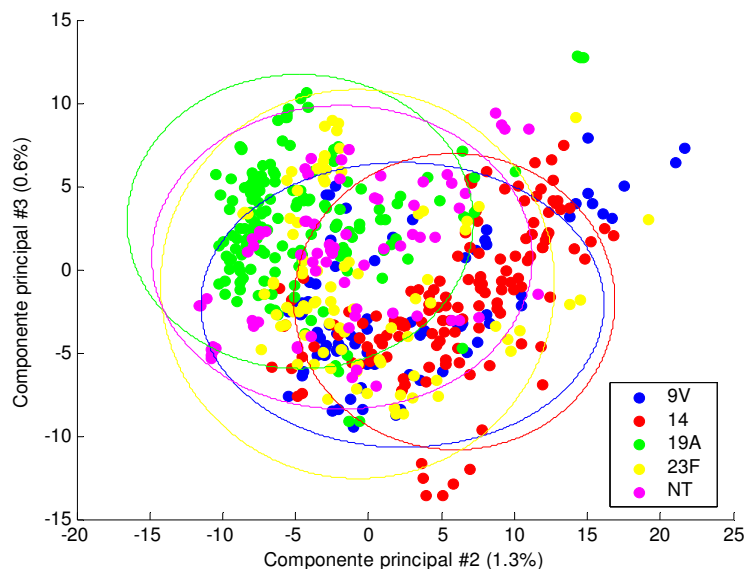


Figura III.12 – Gráfico de componentes principais para todos os espectros analisados, utilizando todos os números de onda, com um pré-processamento de filtro de Savitzky-Golay (9,2,2) e SNV. As 5 cores representam os 5 serótipos do *S. pneumoniae* analisados. A correspondência entre as cores e os serótipos encontra-se na legenda da figura.

Os serótipos dependem dos polissacáridos da parede. Selecionando apenas o intervalo de números de onda que está associado à parede celular ($1185\text{--}900\text{ cm}^{-1}$), aplicando como pré-processamentos o filtro de Savitzky-Golay (9,2,2), e a função SNV para 3 componentes principais, a representação gráfica que se obtém é a que está representada na Figura III.13. Dada a dificuldade em classificar as estirpes segundo o serótipo, houve necessidade de recorrer ao terceiro componente principal, apesar de este representar uma variância explicada pequena. Optou-se por representar o segundo componente principal *versus* o terceiro componente principal, pois verificou-se a existência de uma melhor separação dos serótipos, do que usando o primeiro componente principal.

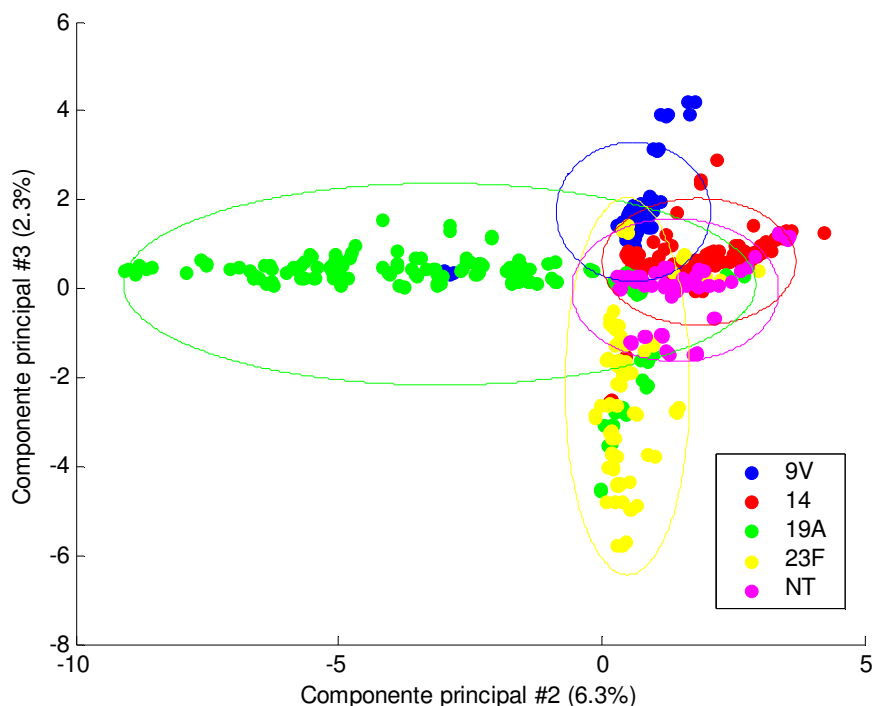


Figura III.13 – Análise de PCA para 542 espectros de estirpes de *S. pneumoniae*. Os espectros usados foram submetidos a um pré-processamento de filtro de Savitzky-Golay (9,2,2) e SNV. A zona do espectro usada correspondente aos números de onda compreendidos no intervalo de 900 a 1185 cm^{-1} .

Na representação do componente principal 3 em função do componente principal 2 é possível distinguir 5 grupos correspondentes aos 5 serótipos. No entanto, esta representação está associada a uma variância explicada baixa. Este valor baixo não quer dizer que seja uma má classificação, apenas indica que a diferença entre os serótipos está contida num componente principal que não representa as características mais marcantes do espectro das células. Na Figura III.13, os intervalos de confiança dos vários serótipos encontram-se alargados devido ao afastamento de alguns pontos. Assim, foi avaliada a existência de espectros aberrantes. Para isso recorreu-se à representação gráfica das estatísticas dos resíduos e Hotelling.

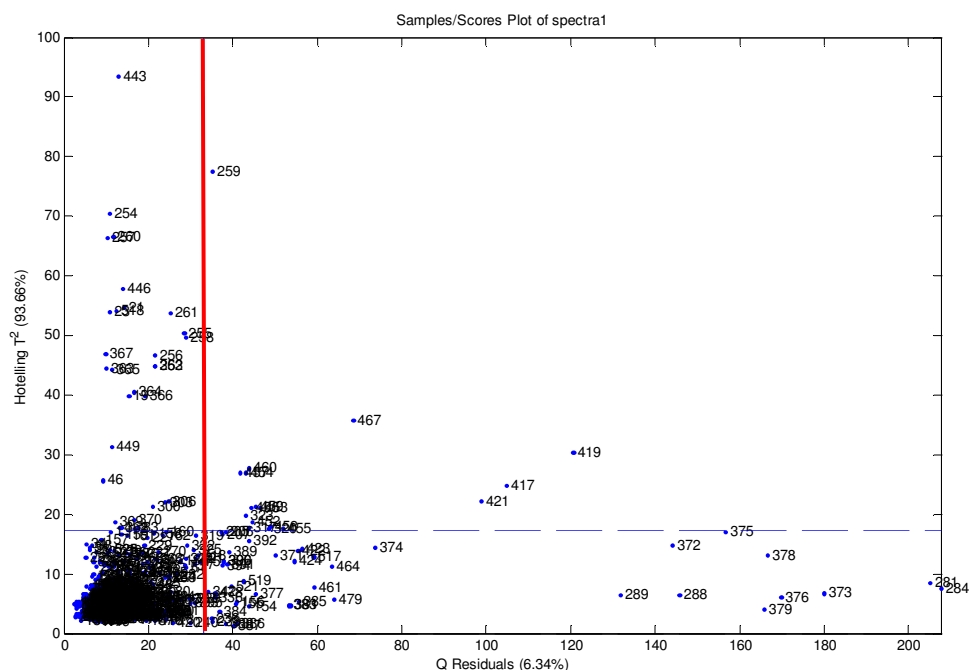


Figura III.14 – Representação gráfica da estatística de Hotelling em função da estatística dos resíduos dos espectros de FT-IR para as estirpes de *S. pneumoniae*, de modo a nomear os espectros aberrantes.

Neste gráfico estão representadas as linhas associadas a 95% de grau de confiança (a encarnado para a estatística dos resíduos e a azul para a estatística de Hotelling). Os espectros que se encontram fora do intervalo de confiança para a estatística de Hotelling (acima da linha azul) e para a estatística dos resíduos (para a direita da linha encarnada) são espectros aberrantes. Assim, a partir da Figura III.15, nomearam-se os seguintes espectros como aberrantes:

- 19, 21, 23 correspondentes à estirpe 1999V0975\$su23F (réplicas de um dia de análise);
- 253, 254, 255, 256, 257, 258, 259, 260, 261 correspondentes à estirpe 2002V1463\$suNT (as 9 réplicas dos 3 dias de análise);
- 281, 284, 288, 289 correspondentes à estirpe 2003V0076\$su19A (réplicas de um dia de análise);

- 317, 323 correspondentes à estirpe 2003V0319\$suNT (ambos do mesmo dia de análise);
- 362, 363, 364, 365, 366, 367 correspondentes à estirpe 2003V0785\$su9V (6 das 9 réplicas correspondentes a 2 dias de análise);
- 372, 373, 375, 376, 378, 379 correspondentes à estirpe 2003V0787\$suNT (6 das 9 réplicas de 2 dias de análise);
- 417, 419, 421 correspondentes à estirpe 2003V1112\$su19A (réplicas de um dia de análise);
- 443, 446 correspondentes à estirpe 2003V1200\$su14 (ambos do mesmo dia de análise);
- 452, 453, 454, 455, 456, 457, 458, 459, 460 correspondentes à estirpe 2003V1203\$suNT (as 9 réplicas dos 3 dias de análise);
- 467, correspondente à estirpe 2003V1481\$su23F ;
- 518, correspondente à estirpe 2003V1494\$su14.

Os espectros considerados aberrantes pela estatística de Hotelling e dos resíduos correspondem a 8% da totalidade dos espectros. A maioria destes espectros corresponde a réplicas de um mesmo dia, o que leva a pensar que tenha havido um erro na preparação da estirpe, nesse dia. Para os espectros cujas réplicas de 2 ou 3 dias foram considerados aberrantes deverá haver uma característica da estirpe que origina espectros não comparáveis com os restantes. A maioria das estirpes (3 de 4) com todos, ou quase todos, os espectros considerados aberrantes pertencem ao grupo NT. A designação de NT corresponde a estirpes em que o serótipo não foi identificado pelo

“método de xadrez” por não haver quantidade de polissacáridos capsulares suficientes ou, por o serótipo adequado não estar representado no método usado. Nestes espectros aberrantes encontram-se representados espectros de 4 das 8 estirpes NT.

Retiraram-se, então, os espectros aberrantes da matriz original e refez-se a representação gráfica (Figura III.15).

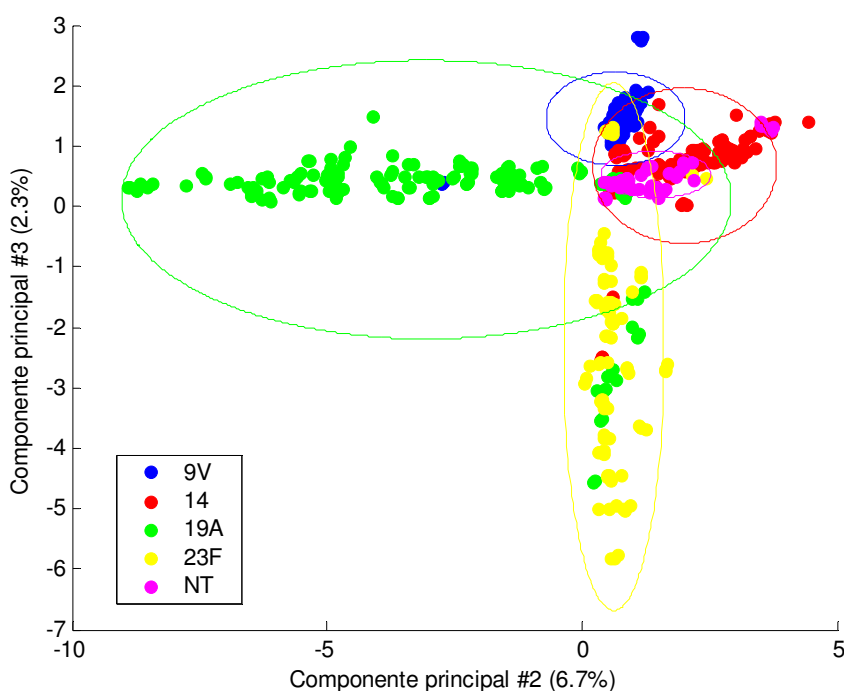


Figura III.15 - Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes de acordo com a representação gráfica da estatística dos resíduos e de Hotelling. Os espectros usados foram submetidos ao pré-processamento filtro de Savitzky-Golay (9,2,2) e SNV, e a gama de números de onda usada foi de 900 a 1185 cm^{-1} . Os serótipos estão organizados por cores e a correspondência encontra-se no gráfico.

Na Figura III.15, pode constatar-se que os intervalos de confiança estão mais apertados que os da Figura III.13. Uma vez que o intervalo de confiança das estirpes com o serótipo não determinado (NT) se encontra sobreposto ao intervalo de confiança dos restantes serótipos, e por não ser um grupo de referência pela sua heterogeneidade, eliminaram-se todas as estirpes

nesta situação. A representação gráfica do resultado obtido encontra-se no gráfico da Figura III.16.

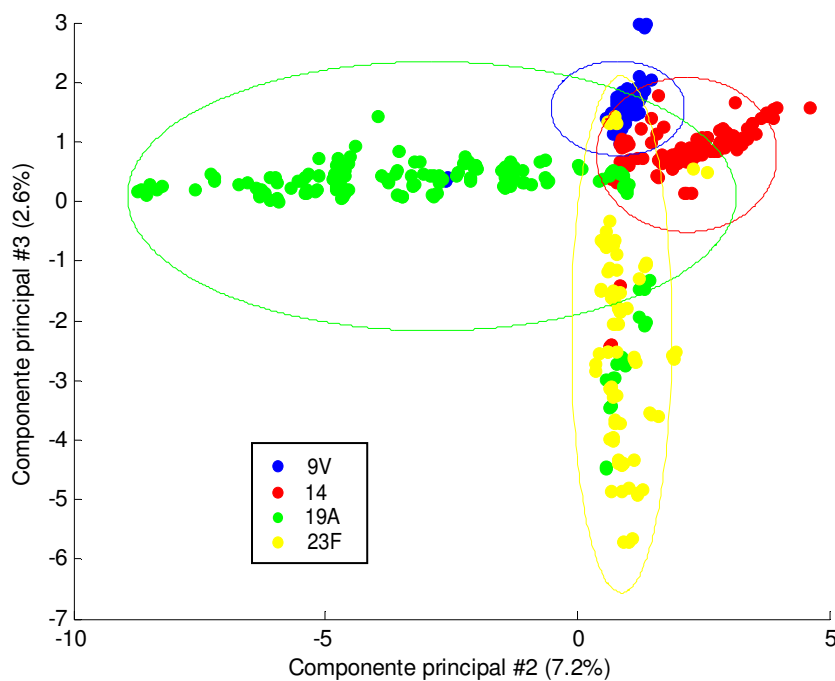


Figura III.16 - Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes e os espectros cujo serótipo não estava determinado pelo método de referência. Os espectros usados foram submetidos aos pré-processamentos filtro de Savitzky-Golay (9,2,2) e SNV, e a gama de números de onda usada foi de 900 a 1185 cm^{-1} . Os serótipos estão organizados por cores e a correspondência encontra-se no gráfico.

Neste gráfico ainda se conseguem detectar alguns espectros de um dado serótipo dentro do intervalo de confiança de outro serótipo. Isto acontece para os casos seguintes:

- Espectros de serótipo 9V mas que se encontram no intervalo de confiança do serótipo 23F:

Dois espectros de um dia de análise e um de outro dia da estirpe 2000V1226\$u9V;

- Espectros de serótipo 19A mas que se encontram no intervalo de confiança do serótipo 23F:

Os 3 espectros de um dia de análise da estirpe 2003V0076\$19A;

Os 3 espectros de um dia de análise da estirpe 2003V0577\$19A;

Todos os espectros da estirpe 2003V1484\$19A;

- Espectros do serótipo 14 mas que se encontram no intervalo de confiança do serótipo 23F:

Os 3 espectros de um dia de análise e 2 espectros de outro dia de análise da estirpe 2000V1210\$23F;

- Espectros de serótipo 23F mas que se encontram no intervalo de confiança do serótipo 14:

Os 2 espectros do mesmo dia de análise do espectro da estirpe 2003V1481\$23F que foi considerado aberrante;

- Espectros de serótipo 23F mas que se encontram no intervalo de confiança do serótipo 9V:

Os 6 espectros de dois dias de análise da estirpe 2000V1217\$23F.

De todos os espectros nomeados, pode observar-se que para a estirpe 2003V1484\$19A todos os espectros estão dentro do intervalo de confiança do serótipo 23F. Uma vez que as análises partiram de placas de gelose de sangue, independentes de dia para dia, e que foram sempre preparadas a partir de um criotubo, o erro não deverá estar no ensaio, mas antes, terá resultado da troca no congelamento da estirpe, uma vez que para a criopreservação foi retirado todo o conteúdo da placa. Como justificação para este caso poderá colocar-se a hipótese de uma troca aquando da repicagem inicial no laboratório de microbiologia da Faculdade de Medicina de Lisboa. Assim, esta estirpe não foi considerada na construção dos modelos. Os restantes espectros que estão no intervalo de confiança de outro serótipo estão

dispostos em múltiplos de três, onde os três são as réplicas de um mesmo dia. Para estes casos, pode colocar-se a hipótese de que tenha havido uma troca de estirpes mas, por não haver certeza quanto a este facto, os espectros não foram eliminados aquando da construção dos modelos. Apenas para perceber a influência que estes espectros que se encontram no intervalo de confiança de outro serótipo têm na representação gráfica da análise de componentes principais, fez-se um novo gráfico onde se eliminaram todos estes espectros. O resultado que se obteve está representado na Figura III.17.

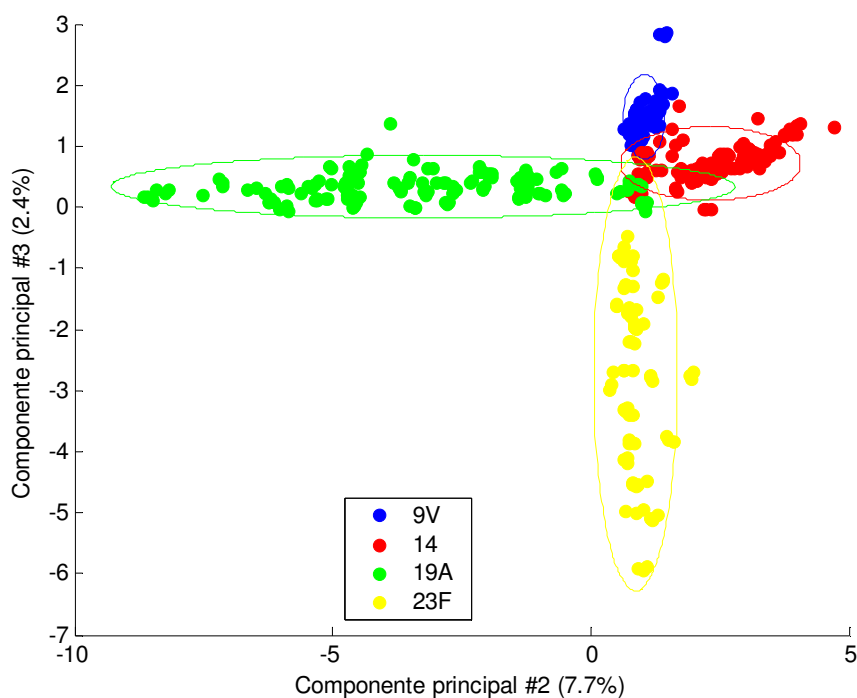


Figura III.17 – Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros de serótipo não determinado e os espectros cujo serótipo se encontrava no intervalo de confiança de outro serótipo. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Os serótipos estão organizados por cores e a correspondência encontra-se na legenda do gráfico.

Como se pode observar no gráfico de Figura III.17, os intervalos de confiança dos serótipos estão bastante apertados. No entanto, pode observar-se ainda uma sobreposição de serótipos 14, 19A e 23F. Se se visualizarem os quatro serótipos num gráfico a três dimensões (Figura III.18) pode-se constatar

que o serótipo 23F tem uma maior variação no componente principal número 3, que o serótipo 19A tem uma maior variação no componente principal número 2 e que o serótipo 9V e 14 têm uma maior variação no componente 1 e estão bem separados entre si.

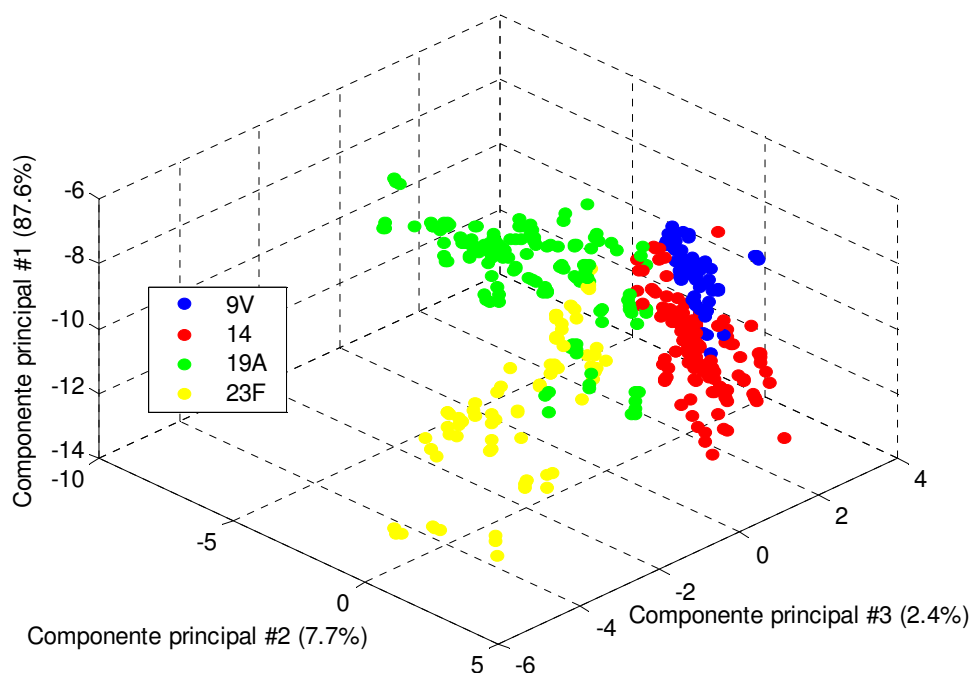


Figura III.18- Representação gráfica da análise de PCA, a 3 dimensões dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros de serótipo não determinado e os espectros cujo serótipo se encontrava no intervalo de confiança de outro serótipo. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Os serótipos estão organizados por cores e a correspondência encontra-se na legenda do gráfico.

Tendo em conta o descrito anteriormente, para efeitos de construção dos modelos nas secções seguintes, consideraram-se como dados de partida o conjunto de espectros resultantes da eliminação dos espectros aberrantes, dos espectros da estirpe 2003V1484\$su19A e dos espectros correspondentes ao serótipo não determinado.

Na abordagem descrita consideraram-se todos os espectros como se fossem espectros independentes. Existe uma outra abordagem possível onde

se escolhe apenas um espectro de cada estirpe que esteja na nuvem representada pelas 9 réplicas. Assim sendo, obtém-se um conjunto de pontos significativamente mais pequeno. Uma vez que as estirpes cujo serótipo não está determinado têm uma grande variabilidade, foram eliminados logo à partida e fez-se uma representação gráfica (Figura III.19) com os pontos referentes às 52 estirpes restantes.

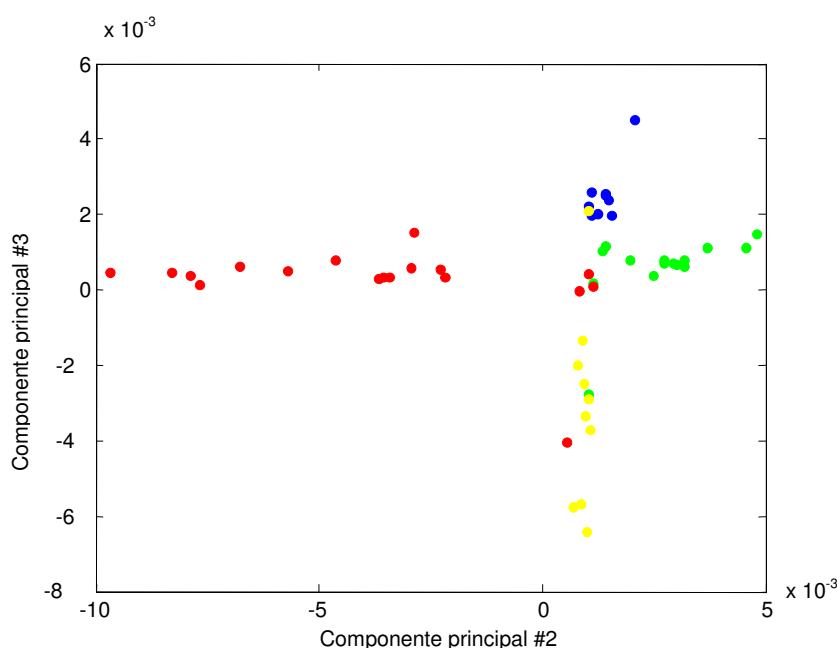


Figura III.19 - Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros de serótipo não determinado. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Cada ponto corresponde a um espectro representativo de uma nuvem de espectros correspondente a uma estirpe. A legenda da figura é a seguinte: (●) serótipo 9V, (●) serótipo 14, (●) serótipo 19A, e (●) serótipo 23F.

Como se pode ver na Figura III.19, existe uma dispersão pronunciada nos componentes da análise de PCA, na Figura III.20 identificam-se visualmente estas estirpes que perfazem um total de 11.

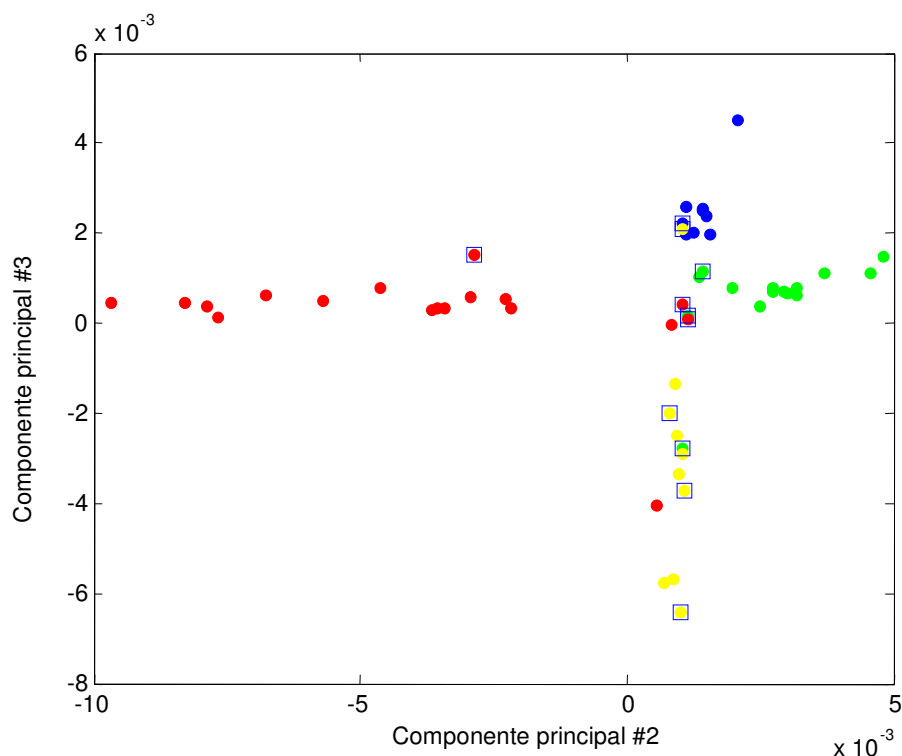


Figura III.20 - Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros de serótipo não determinado. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Cada ponto corresponde a um espectro representativo de uma nuvem de espectros correspondente a uma estirpe. Com um quadrado estão representadas as estirpes com uma grande dispersão. A legenda da figura é a seguinte: (●) serótipo 9V, (●) serótipo 14, (●) serótipo 19A, e (●) serótipo 23F.

Retirando as 11 estirpes restou com um conjunto de 41 estirpes.

Construindo um novo modelo PCA obtém-se a representação gráfica na Figura III.21.

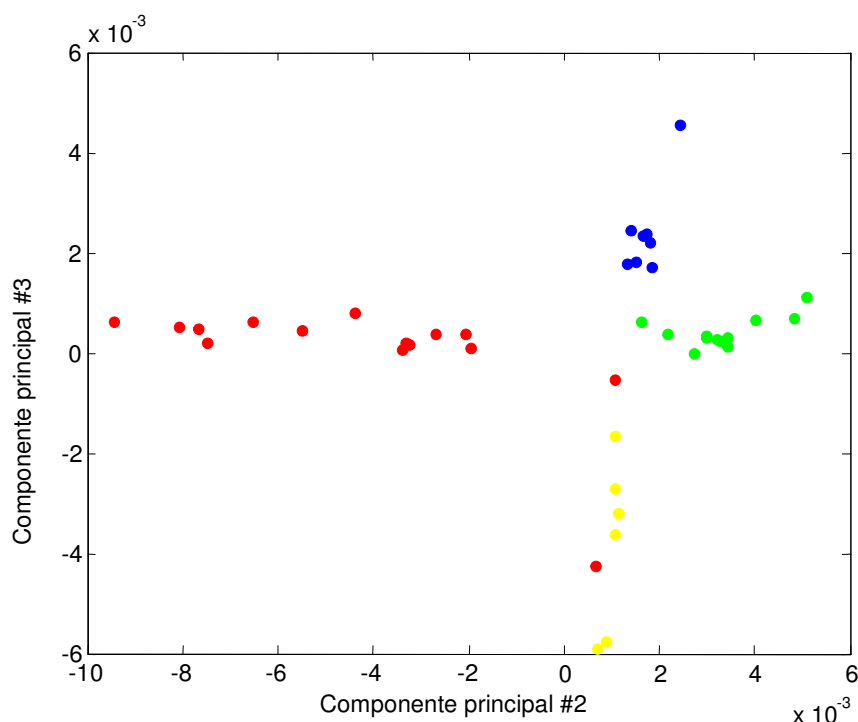


Figura III.21 - Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros de serótipo não determinado e os espectros que apresentavam uma grande dispersão. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Cada ponto corresponde a um espectro representativo de uma nuvem de espectros correspondente a uma estirpe. A legenda da figura é a seguinte: (●) serótipo 9V, (●) serótipo 14, (●) serótipo 19A, e (●) serótipo 23F.

Retirando a estirpe 2003V1484\$u19A, a estirpe cujos espectros se encontram todos no intervalo de confiança do serótipo 23F, obtém-se a representação gráfica da Figura III.22.

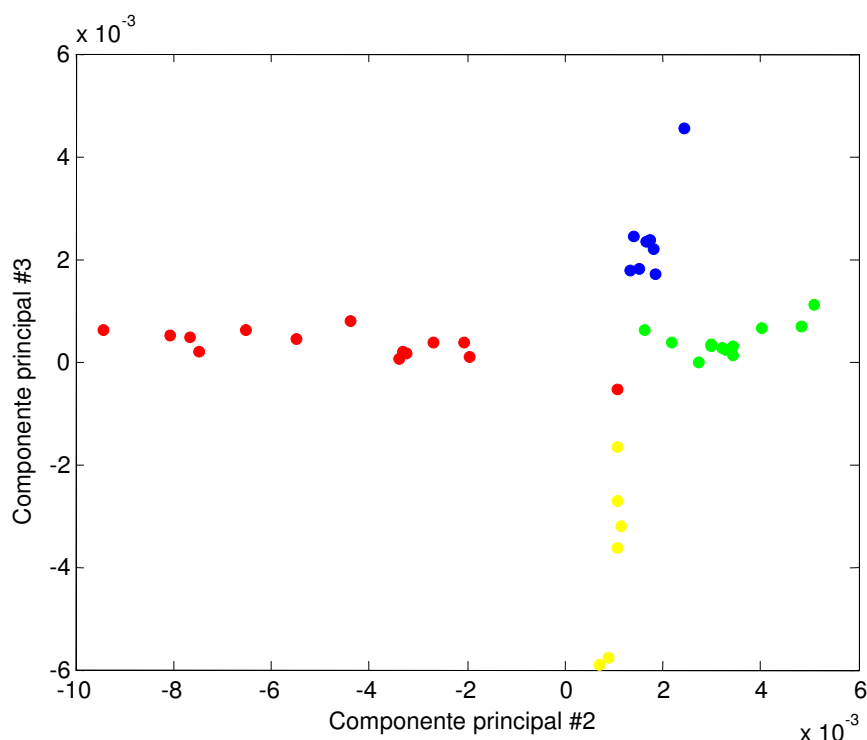


Figura III.22 - Representação gráfica da análise de PCA dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminadas as estirpes de serótipo não determinado, as estirpes que apresentavam uma grande dispersão e a estirpe 2003V1484\$su19A. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Cada ponto corresponde a um espectro representativo de uma nuvem de espectros correspondente a uma estirpe. A legenda da figura é a seguinte: (●) serótipo 9V, (●) serótipo 14, (●) serótipo 19A, e (●) serótipo 23F.

Das duas abordagens acima descritas, seguiu-se a primeira para a construção dos modelos, uma vez que esta traduz a variabilidade inerente ao crescimento das estirpes nos três lotes de meio, à preparação da amostra, à secagem das suspensões no suporte de Selenato de Zinco e à própria leitura dos resultados. A metodologia seguida foi sempre a mesma mas, apesar de todos os controlos, existe sempre alguma variabilidade que pode ser visível nos resultados. A segunda abordagem permite diminuir a incerteza associada à determinação de um serótipo por este método mas, por outro lado, como o intervalo de confiança está mais apertado poderá dar origem a falsos negativos para um dado serótipo. Assim, a abordagem associada a uma maior reprodutibilidade será a primeira.

Para a elaboração dos modelos, é necessária a divisão dos espectros de modo a que existam espectros para a construção do modelo e espectros distintos para a sua validação. Assim, dividiu-se a totalidade dos espectros em dois grupos de modo a que um dos grupos contenha 70% dos espectros de cada serótipo e o outro contenha os 30% dos espectros restantes de cada serótipo. De modo a avaliar se os dados de validação escolhidos são representativos de cada serótipo fez-se uma representação gráfica de cada serótipo onde se colocaram os espectros pertencentes ao grupo de validação (pontos a preto) sobrepostos ao grupo de espectros usados para construção do modelo (pontos a cinzento). Os pontos correspondentes aos dados de validação deverão estar contidos nos dados de construção do modelo, para uma divisão bem feita. A representação gráfica para os 4 serótipos encontra-se na Figura III.23.

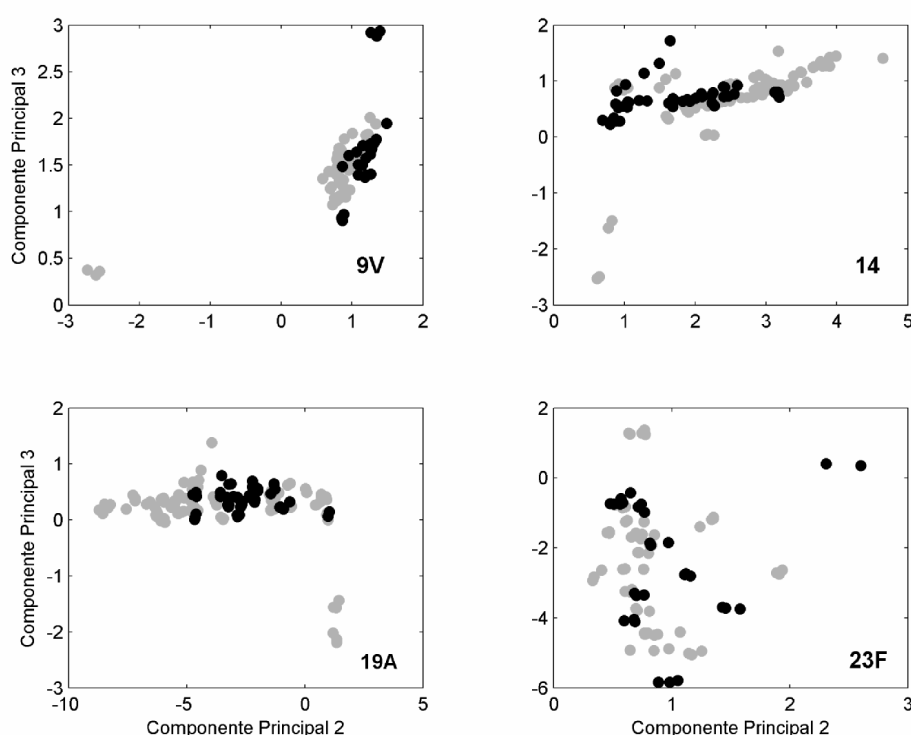


Figura III.23— Representação gráfica dos espectros usados para a validação dos modelos (a cinzento escuro) e os espectros usados na construção dos modelos (a cinzento claro), para cada serótipo (9V, 14, 19A e 23F).

Para todos os gráficos dos serótipos 9V, 14 e 19A os dados de validação estão contidos nos de construção do modelo, mas para o serótipo 23F existem 2 espectros ligeiramente afastados. Este serótipo, como foi visto na análise de PCA, tem uma grande variação no componente principal 3 o que justifica a dispersão dos dados e não existe razão para eliminar os dois espectros que parecem estar mais afastados do grupo considerado.

5.1.2. SIMCA

A análise de componentes principais por si só não permite a identificação de estirpes de amostras desconhecidas. Assim, com as informações obtidas aquando da análise de componentes principais, pode-se construir um modelo utilizando um método supervisionado. Um dos métodos supervisionados possíveis é o SIMCA. Para a construção do modelo SIMCA tem de se criar uma matriz numérica onde cada número corresponde a um serótipo. Esta matriz tem apenas uma coluna e está organizada de modo a que cada linha corresponda ao serótipo de cada espectro da matriz dos espectros, codificada de 1 a 5. No modelo SIMCA é, também, necessário o cálculo do número de componentes principais adequado, cálculo este que foi feito por validação cruzada. Na validação cruzada teve-se o cuidado para que não se validasse o modelo usando estirpes também usadas na calibração. Para isso não acontecer usou-se a validação cruzada através da eliminação de blocos de exemplos (LOBO). Na validação cruzada, usando como matriz analisada aquela que contém a totalidade dos espectros de FT-IR (542 espectros), obteve-se uma representação gráfica que indica que o número de componentes principais associado a um menor erro é 2. Assim, fazendo o

modelo SIMCA com 2 componentes principais, o que se obtém é o representado na Figura III.24.

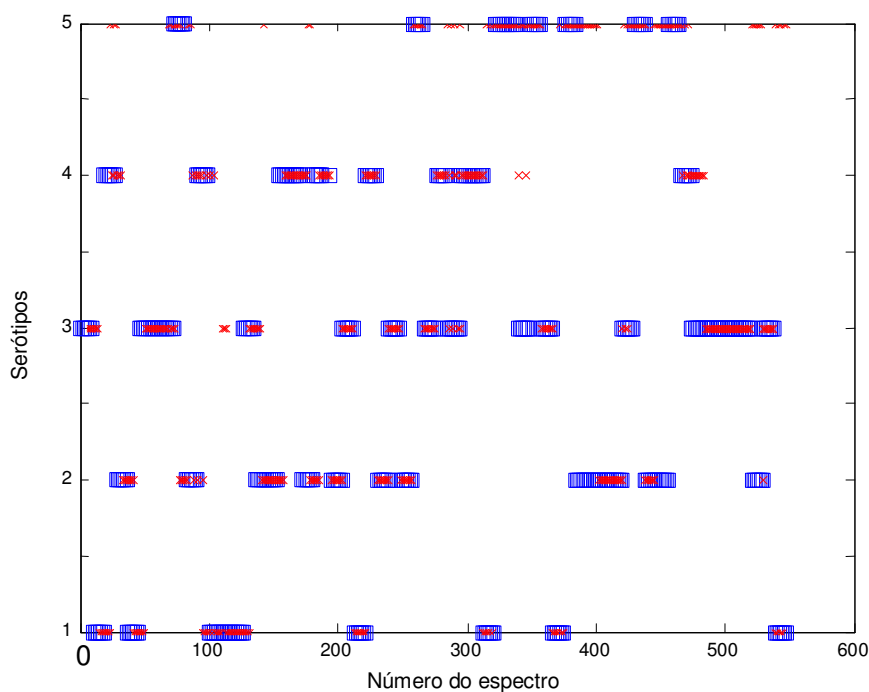


Figura III.24 – Resultado da validação cruzada para o modelo SIMCA com 2 componentes principais para o conjunto total de espectros e para 16 blocos aleatórios. Os espectros foram submetidos ao pré-processamento filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número 1 corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A, o 4 ao 23F e o 5 ao serótipo não determinado. Representada com o símbolo (x) está a classe prevista pelo modelo e com o símbolo (□) a classe definida pelos métodos clássicos. Existe uma classificação incorrecta para 20,7% dos espectros.

Neste modelo, 20,7% dos espectros foram mal identificados. Tal como já foi referido, este modelo baseia-se na análise de componentes principais e, em semelhança ao que foi feito para o PCA, construiu-se um novo modelo SIMCA para o conjunto de espectros resultantes da eliminação dos espectros aberrantes, segundo a estatística de Hotelling e dos resíduos, e sem a estirpe 2003V1484\$su19A. A Figura III.25 representa esse estudo.

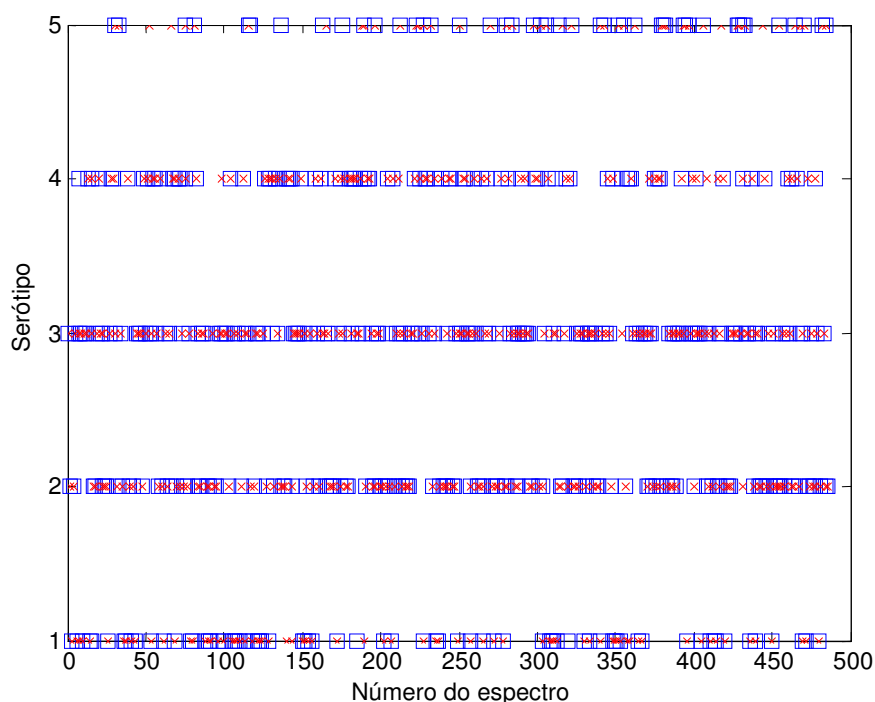


Figura III.25 – Validação cruzada para o modelo SIMCA para 16 blocos aleatórios com 4 componentes principais para o conjunto total de espectros dos quais foram eliminados os espectros aberrantes e os espectros da estirpe 2003V1484\$u19A. Os espectros foram submetidos ao pré-processamento filtro de Savitzky-Golay (9,2,2) e SNV, e foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número 1 corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A, o 4 ao 23F e o 5 ao serótipo não determinado. Representada com o símbolo (x) está a classe prevista pelo modelo e com o símbolo (□) a classe definida pelo método clássico. Existe uma classificação errada para 8,2% dos espectros.

A percentagem de classificação dos serótipos correcta (91,8%) aumentou com a eliminação dos espectros aberrantes e dos espectros da estirpe 2003V1484\$u19A. No entanto, ainda existem 40 espectros que não foram correctamente identificados (correspondentes a 8,2%). A maioria destes espectros encontra-se na classe 5, ou seja, nos espectros de estirpes cujo serótipo não foi determinado (NT). Assim, se se eliminarem estas estirpes o resultado do modelo SIMCA resultante está descrito na Figura III.26.

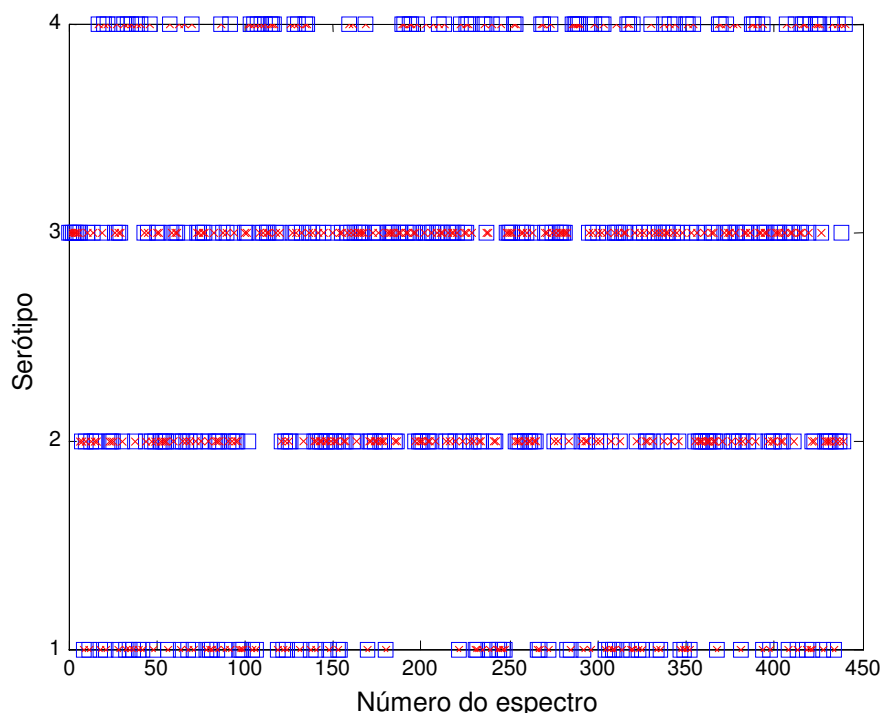


Figura III.26 – Validação cruzada para 16 blocos aleatórios para o modelo SIMCA com 3 componentes principais para o conjunto total de espectros dos quais foram eliminados os espectros aberrantes, os espectros da estirpe 2003V1484\$su19A e os espectros sem serótipo determinado. Os espectros foram submetidos ao pré-processamento filtro de Savitzky-Golay (9,2,2) e SNV e, foi usada a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número 1 corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representada com o símbolo (x) está a classe prevista pelo modelo e com o símbolo (□) a classe definida pelo “método de xadrez”. Existe uma classificação incorrecta para 5,5% das estirpes.

O modelo SIMCA assim construído conseguiu prever uma classificação concordante com o “método de xadrez” para 94,5% dos espectros. Assim, deve dividir-se a matriz dos espectros de modo a ter dados para a construção do modelo e dados para testar o modelo, que pertençam a estirpes distintas. Usando os 70% dos espectros, referidos anteriormente (secção 5.1.1), para a construção do modelo e os 30% dos espectros para a sua validação, e após a determinação do número de componentes principais associado a um menor erro para 16 blocos aleatórios, foi elaborado um novo modelo SIMCA com 2 componentes principais. A Figura III.27 mostra este modelo.

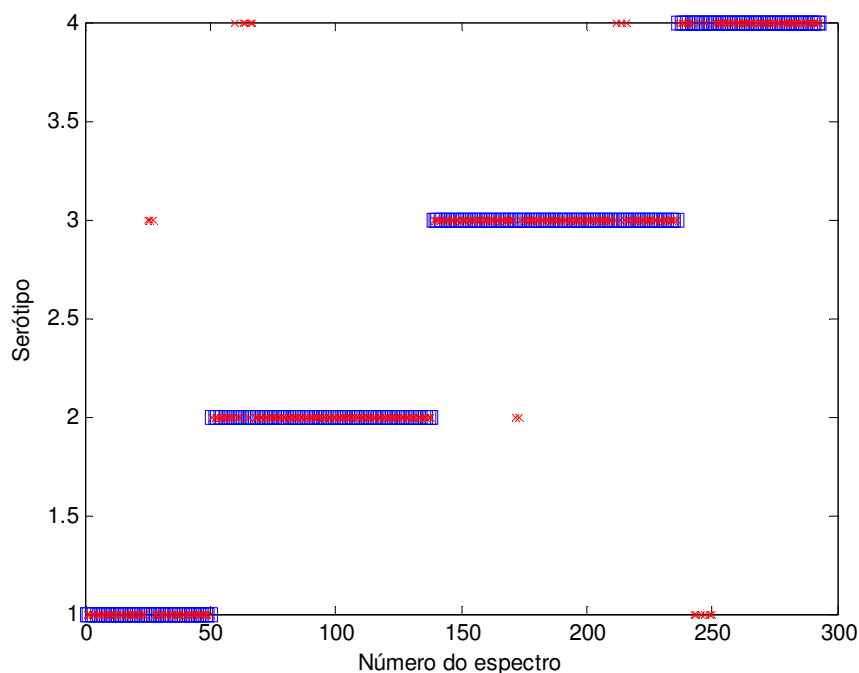


Figura III.27 – Modelo SIMCA para 2 componentes principais para 70% dos espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros da estirpe 2003V1484\$su19A e os espectros sem serótipo determinado. Os espectros usados foram submetidos aos pré-processamentos: Savitzky-Golay (9,2,2) e SNV e usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número 1 corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representado com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo teste de referência. Existe uma classificação correcta para 93,5% dos espectros.

A percentagem de espectros correctamente associados ao serótipo é bastante elevada (93,5%) no entanto, é necessário testar o modelo com os 30% do total dos espectros (133 espectros) que não foram usados na construção do modelo. O resultado obtido encontra-se na Figura III.28.

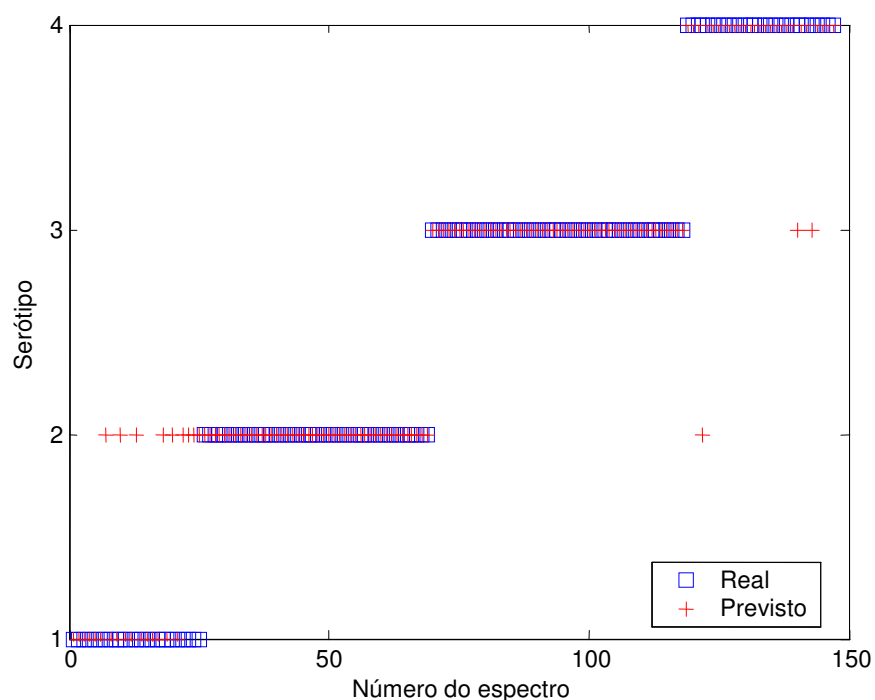


Figura III.28 – Validação do Modelo SIMCA para 30% da totalidade dos espectros de FT-IR de *S. pneumoniae* dos quais foram eliminados os espectros aberrantes e os espectros sem serótipo determinado. Os espectros usados foram submetidos aos pré-processamentos filtro de Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número 1 corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representada com o símbolo (+) está a classe prevista pelo modelo e com o símbolo (□) a classe definida previamente. A percentagem de espectros bem classificados foi de 91,8.

Tanto o modelo como a validação apresentaram uma percentagem de estirpes bem classificadas muito elevada (93,5% e 91,8%, respectivamente), sendo, por isso, uma abordagem que pode ser considerada para a identificação de um dos 4 serótipos testados.

5.1.3. PLS

Um outro modelo supervisionado é a regressão dos mínimos quadrados parciais (PLS). Este modelo pode ser aplicado de duas formas: PLS discriminante e PLS hierárquico.

5.1.3.1. PLS Discriminante

No PLS discriminante são consideradas duas matrizes: uma matriz X , que representa a variável independente que contém os espectros de FT-IR a analisar, e uma matriz Y que representa a variável dependente que tem a informação sobre o serótipo de cada espectro. A matriz Y tem um número de linhas igual ao número de linhas da matriz X , uma vez que as linhas correspondem aos espectros a estudar, e um número de colunas igual ao número de serótipos. Na matriz Y só existem números “1” e “0”, e se se definir como a primeira coluna o serótipo 9V, a segunda como o serótipo 14, a terceira como o serótipo 19A, a quarta como o serótipo 23F e a quinta como o serótipo não determinado, para cada linha o número “1” é colocado na coluna referente ao serótipo adequado, as restantes colunas que dizem respeito aos outros serótipos são preenchidas com o número 0. Tal como para o método PCA, os espectros devem ser submetidos a um pré-processamento adequado. De acordo com a informação obtida pelo métodos PCA e SIMCA, usou-se como pré-processamento o filtro de Savitsky-Golay (9,2,2) e a função SNV. Também foi usada a gama espectral correspondente à parede celular: 900 a 1185 cm^{-1} . Para a aplicação desta função deve definir-se o número de variáveis latentes que foi determinado por validação cruzada para 10 blocos contíguos. O erro mínimo foi obtido para um número de variáveis latentes de 5. Aplicando então esta função à totalidade dos espectros, a representação gráfica que se obteve é a representada na Figura III.29.

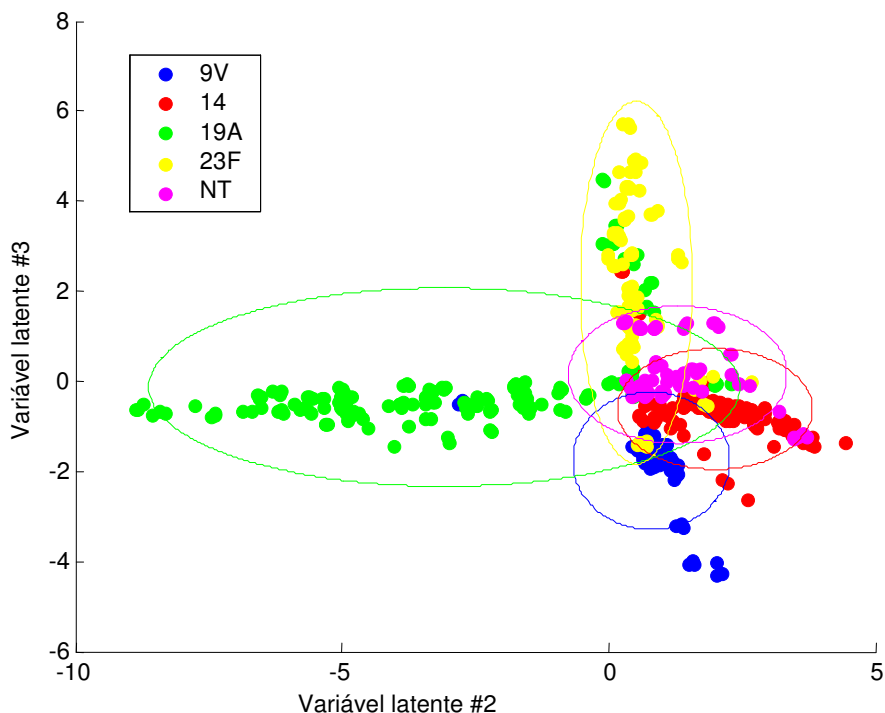


Figura III.29 - Resultados obtidos para o modelo de PLS discriminante usando para a validação cruzada 10 blocos contíguos e 5 variáveis latentes. Foram usados todos os espectros FT-IR de *S. pneumoniae*, que foram submetidos a pré-processamentos Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . A equivalência entre as cores e os serótipos encontra-se na legenda.

Esta representação é muito semelhante à representação gráfica do método PCA, para este mesmo conjunto de espectros. Assim, tal como foi feito para o PCA e para o SIMCA, eliminaram-se os espectros aberrantes e os espectros da estirpe 2003V1484\$su19A, a representação gráfica que se obtém é a representada na Figura III.30.

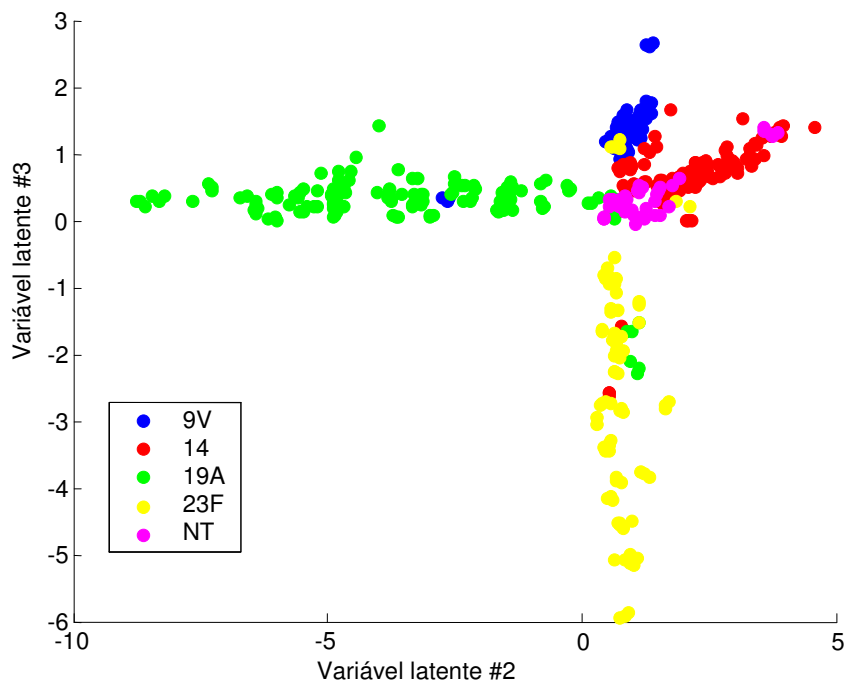


Figura III.30 - Resultados obtidos para o modelo de PLS discriminante para uma validação cruzada com 10 blocos contíguos. Foi usado um conjunto de espectros FT-IR de *S. pneumoniae* que resultam da eliminação dos espectros aberrantes de acordo com o descrito em 3.1.2.1. Os espectros foram submetidos aos pré-processamentos filtro de Savitzky-Golay (9,2,2) e SNV, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . A equivalência entre as cores e os serótipos encontra-se na legenda.

Os espectros correspondentes às estirpes com uma quantidade pequena de polissacáridos da parede ou com um serótipo não incluído no “método de xadrez”, apresentam-se sobrepostos aos dos 4 serótipos testados. Como este grupo poderá ter uma variabilidade muito grande, os espectros correspondentes foram eliminados e foram testados apenas os 4 serótipos (9V, 14, 19A, 23F). O resultado é o que está representado na Figura III.31.

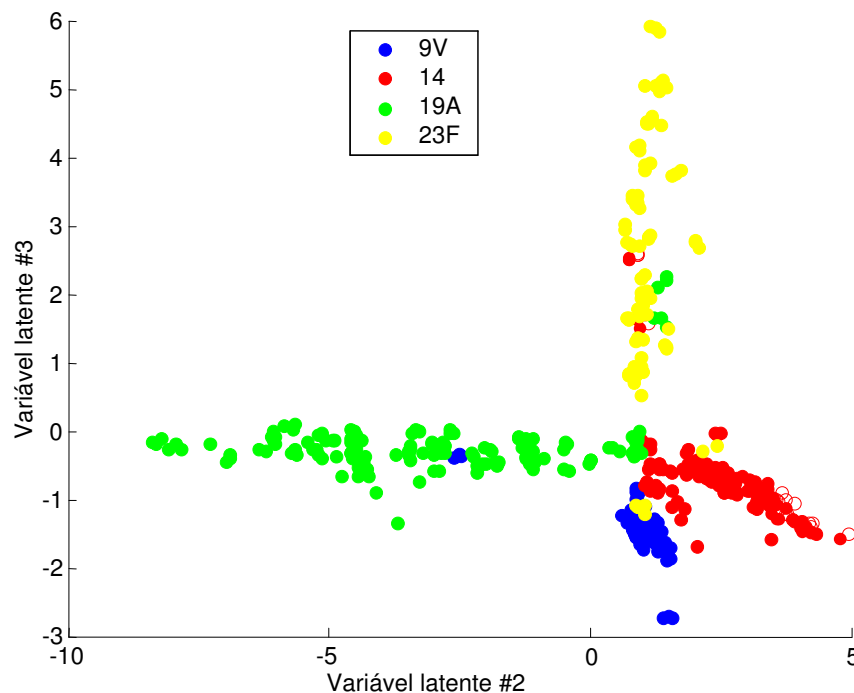


Figura III.31 - Resultados obtidos para o modelo de PLS discriminante para o conjunto de espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros da estirpe 2003V1484\$u19A e os espectros cujo serótipo não está determinado. Este modelo resulta de uma validação cruzada com 10 blocos contíguos e 5 variáveis latentes. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . A equivalência entre as cores e os serótipos encontra-se na legenda.

Como se pode ver na Figura III.31, os 4 serótipos apresentam alguma separação mas, existe um ponto coincidente. Usando os 70% dos espectros, referidos no ponto 5.1.1, para a construção do modelo PLS discriminante, o que se obteve é o representado na Figura III.32.

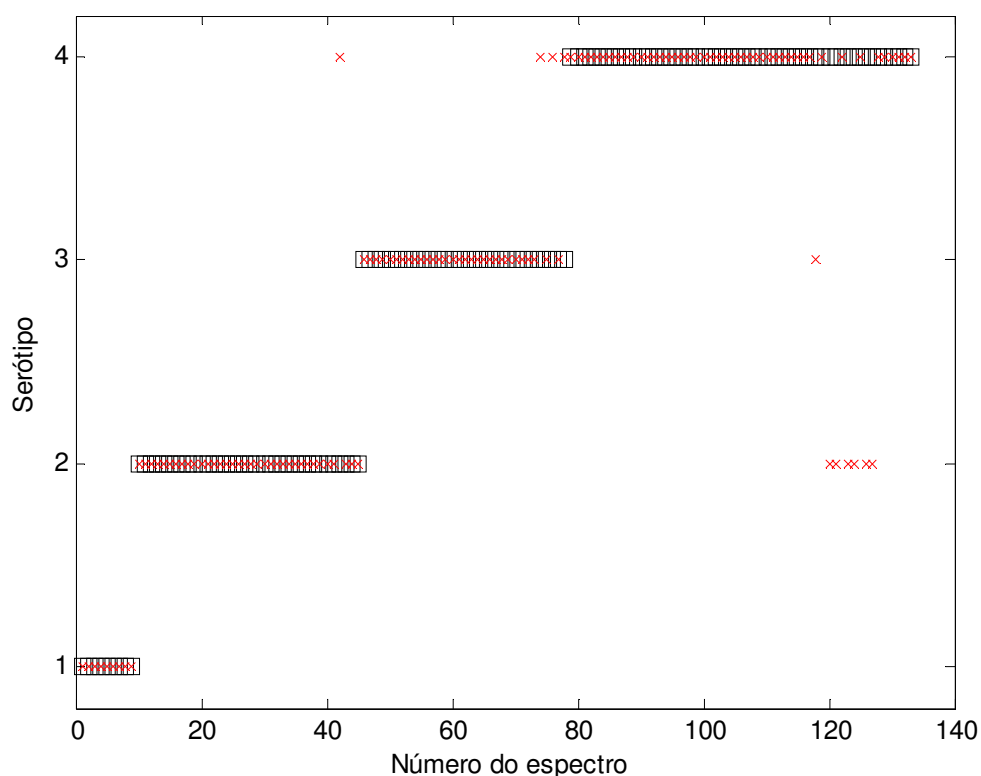


Figura III.32 - Resultados obtidos para o modelo de PLS discriminante para o conjunto de espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros cujo serótipo não está determinado e os espectros da estirpe 2003V1484\$u19A. Para a construção do modelo foram usadas 5 variáveis latentes. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”. Com a designação 1 está o serótipo 9V, 2 o serótipo 14, 3 o serótipo 19A e 4 o serótipo 23F. A percentagem de espectros correctamente identificados foi de 92,6%.

O modelo construído testa os próprios dados e, para estes, a percentagem de espectros correctamente identificados foi de 92,6%.

De modo a validar o modelo usaram-se os 30% dos espectros que não foram usados na sua construção. O resultado que se obtém para a validação é o representado na Figura III.33.

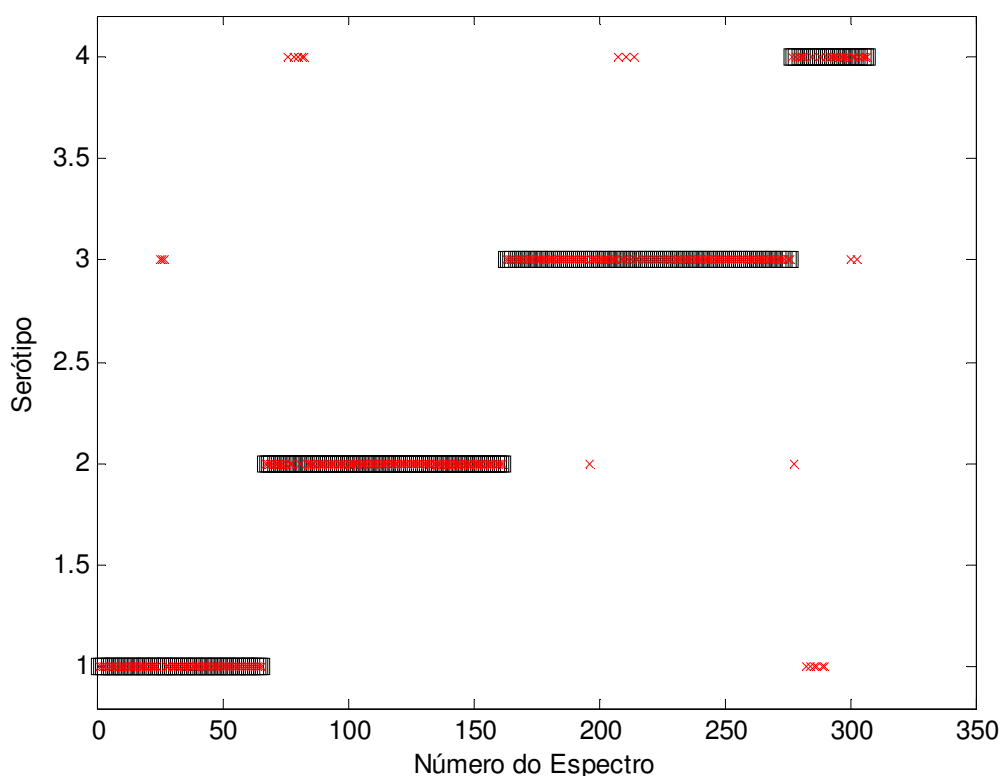


Figura III.33 - Resultados obtidos para a validação do modelo de PLS discriminante para 30% do conjunto de espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes e os espectros cujo serótipo não está determinado. Este modelo resulta de uma validação cruzada com 5 variáveis latentes e com 1000 re-amostragens. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (\square) o serótipo definido pelo “método de xadrez”. Com a designação 1 está o serótipo 9V, 2 o serótipo 14, 3 o serótipo 19A e 4 o serótipo 23F. A percentagem de estirpes correctamente identificadas foi de 97,1%.

A percentagem de concordância entre a determinação dos serótipos pelo método de referência e o do modelo PLS discriminante é de 97,1% o que representa um valor bastante elevado e permite concluir que esta é uma abordagem válida para a serotipagem de estirpes que pertençam a um destes 4 serótipos.

Na Figura III.34 encontram-se quatro gráficos, um por cada serótipo. Em cada gráfico está representada a incerteza associada a cada previsão do serótipo pelo modelo, para um grau de confiança de 95%. Assim por cada serótipo, e consequentemente por cada gráfico, os espectros que se encontram

na posição 1 são aqueles que estão mais próximos do serótipo estudado e na linha correspondente ao zero encontram-se todos os espectros dos três serótipos restantes.

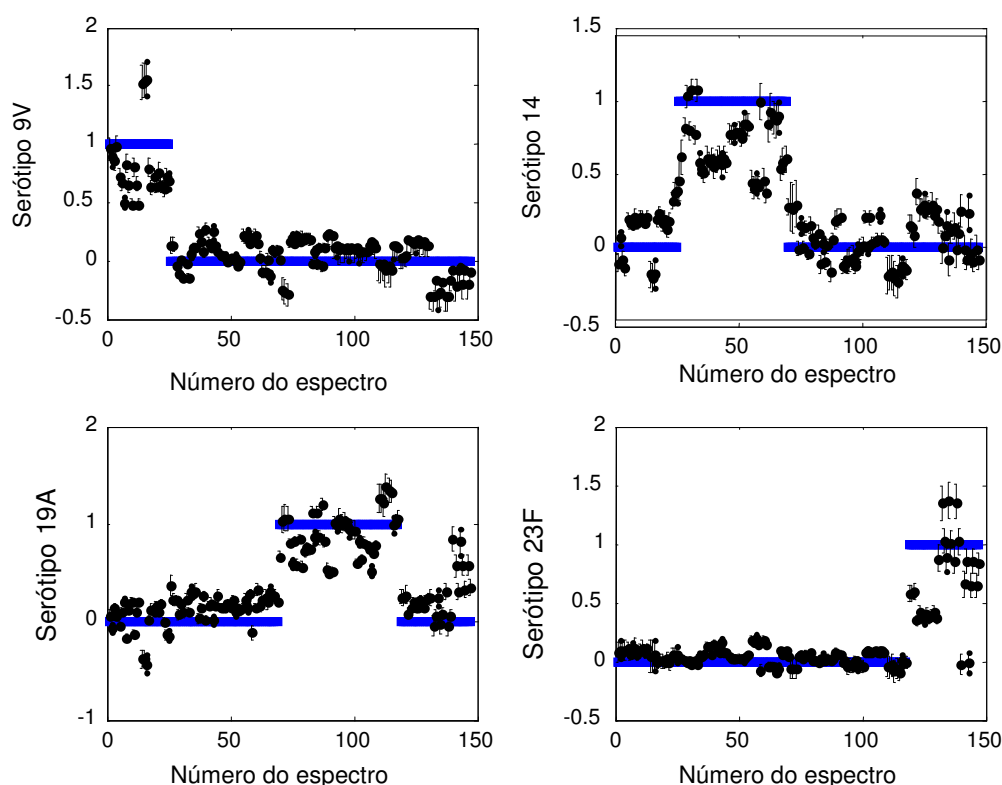


Figura III.34 – Resultados das incertezas para a validação do modelo de PLS discriminante para 30% do conjunto de espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros da estirpe 2003V1484\$su19A e os espectros cujo serótipo não está determinado. Este modelo resulta de uma validação cruzada com 5 variáveis latentes e com 100 re-amostragens. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Em cada gráfico na linha 1 encontram-se os espectros previstos pelo modelo associados ao serótipo respectivo e na linha zero os espectros das outras 3 classes. Os espectros foram ordenados de acordo com o serótipo determinado pelo “método de xadrez”.

De acordo com o representado na Figura III.34, o serótipo cujos espectros têm uma maior incerteza associada é o 23F. Como se pode observar pela representação gráfica do PLS discriminante da Figura III.31, os espectros deste serótipo têm uma variabilidade grande na variável latente 3 e, para além disto, o número de espectros usados para a validação deste serótipo é o mais elevado, o que quer dizer que o número de espectros usados na construção do

modelo foi baixo levando à redução da variabilidade aquando da construção do modelo o que se traduz numa menor reprodutibilidade e, portanto, numa incerteza maior para cada espectro deste serótipo.

De modo a comparar os resultados obtidos pelo modelos PLS discriminante e pelo SIMCA foi feita uma representação gráfica (Figura III.35) de ambos os métodos para vários componentes principais.

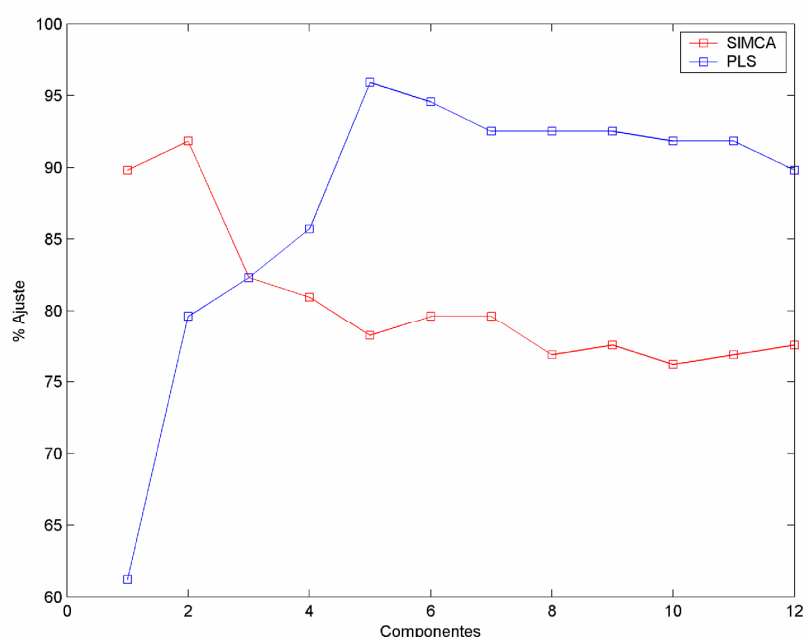


Figura III.35 – Representação gráfica dos resultados obtidos para modelos SIMCA e PLS para o mesmo conjunto de dados para vários componentes principais.

De acordo com a Figura III.35, verifica-se que existe um melhor ajuste para o modelo PLS discriminante quando se considera um número de componentes superior a 4. O modelo SIMCA foi construído para 2 componentes principais e o PLS discriminante para 5 variáveis latentes, os pontos mais elevados para ambos os modelos. No entanto, segundo o gráfico, a percentagem de ajuste é mais elevada para o PLS discriminante como se pode constatar pela percentagens de estirpes correctamente identificadas

aquando da validação dos modelos SIMCA (91,8%) e PLS discriminante (97,1%).

5.1.3.2. PLS hierárquico

Outro modelo supervisionado é o PLS hierárquico. Neste modelo são eliminados inicialmente os espectros do serótipo que melhor se distingue do grupo. Para a correcta aplicação do PLS hierárquico, é realizada uma validação cruzada de modo a determinar o número de variáveis latentes associadas a um erro menor. Do grupo de espectros que resta (equivalentes a 4 serótipos), são eliminados os espectros do serótipo que melhor se distingue do grupo de espectros dos 3 serótipos restantes, sendo novamente determinado o número de variáveis latentes adequado, através de validação cruzada. Assim vão-se eliminando sucessivamente serótipos até que fica a última classe.

Construiu-se um modelo PLS hierárquico para os espectros pré-processados com as funções Savitzky-Golay (9,2,2) e SNV, usando apenas os números de onda entre 900 e 1185 cm^{-1} , definindo 12 blocos contíguos para a validação cruzada. Obteve-se o modelo descrito na Tabela III.3.

Tabela III.3 – Resultados obtidos para o modelo de PLS hierárquico usando para a validação cruzada 12 blocos contíguos. Foram usados todos os espectros FT-IR de *S. pneumoniae*, que foram submetidos a pré-processamentos Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Nesta tabela encontram-se representados os serótipos, o número de variáveis latentes usadas para a identificação de cada serótipo, a sensibilidade e a especificidade.

Serótipo	LV	Sensibilidade (α)	Especificidade (β)
9V	4	96,30	98,70
14	12	94,07	96,32
23F	10	93,33	87,29
19A	15	93,90	91,67
NT	-	100	100

Para se entender a tabela é importante definir a sensibilidade como os espectros correctamente classificados, aos quais foram retirados os falsos

positivos, e a especificidade como os espectros correctamente classificados, aos quais foram retirados os falsos negativos. No modelo ideal dever-se-ia obter, para a sensibilidade e especificidade, para cada serótipo, um valor próximo de 100.

Tal como descrito nas secções anteriores, foram eliminados os espectros aberrantes e os espectros da estirpe 2003V1484\$su19A, e para 12 blocos contíguos na validação cruzada, o modelo que se obteve é o descrito na Tabela III.4.

Tabela III.4 - Resultados obtidos para o modelo PLS hierárquico para uma validação cruzada com 12 blocos contíguos. Foi usado um conjunto de espectros FT-IR de *S. pneumoniae* que resultam da eliminação dos espectros aberrantes pela estatística de Hotelling e dos resíduos e da eliminação dos espectros da estirpe 2003V1484\$su19A. Os espectros foram submetidos aos pré-processamentos filtro de Savitzky-Golay (9,2,2) e SNV, usou-se a gama de números de onda de 900 a 1185 cm⁻¹. Nesta tabela encontram-se representados os serótipos, o número de variáveis latentes usadas para a identificação de cada serótipo, a sensibilidade e a especificidade.

Serótipo	LV	Sensibilidade (α)	Especificidade (β)
9V	3	96,00	98,54
23F	6	97,67	93,85
14	3	100,00	93,78
19A	1	83,67	100,00
NT	-	100,00	100,00

Tal como foi feito nas secções anteriores, retirou-se, ao conjunto de espectros resultantes da eliminação dos espectros aberrantes, os espectros cujo serótipo não foi identificado com sucesso pelo método de referência, e para 12 blocos contíguos para a validação cruzada, o modelo PLS hierárquico obtido é o descrito na Tabela III.5.

Tabela III.5 - Resultados obtidos para o modelo PLS hierárquico para o conjunto de espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes, os espectros da estirpe 2003V1484\$su e os espectros cujo serótipo não está determinado. Este modelo resulta de uma validação cruzada com 12 blocos contíguos. Os espectros usados foram submetidos aos pré-processamentos: filtro de Savitzky-Golay (9,2,2) e SNV e, usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Nesta tabela encontram-se representados os serótipos, o número de variáveis latentes usadas para a identificação de cada serótipo, a sensibilidade e a especificidade.

Classe	LV	Sensibilidade (α)	Especificidade (β)
9V	3	96,00	98,36
19A	4	95,92	99,08
14	3	95,46	91,86
23F	-	100,00	100,00

Houve uma melhoria significativa na sensibilidade e especificidade do modelo obtido. Uma vez que os modelos construídos devem ser testados com dados que não foram usados na elaboração do modelo, construiu-se um novo modelo com 70% dos espectros. O resultado obtido encontra-se na Tabela III.6 e na Figura III.36.

Tabela III.6 - Resultados obtidos para o modelo PLS hierárquico para 70% do conjunto de espectros de FT-IR de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes e os espectros cujo serótipo não foi determinado. Os espectros usados foram submetidos aos pré-processamentos Savitzky-Golay (9,2,2) e SNV e usou-se a gama de números de onda de 900 a 1185 cm^{-1} . O modelo resulta de uma validação cruzada com 12 blocos contíguos. Nesta tabela encontram-se representados os serótipos, o número de variáveis latentes usadas para a identificação de cada serótipo, a sensibilidade e a especificidade.

Classe	LV	Sensibilidade (α)	Especificidade (β)
19A	9	96,94	98,97
9V	1	100,00	95,86
14	4	94,32	100,00
23F	-	100,00	100,0000

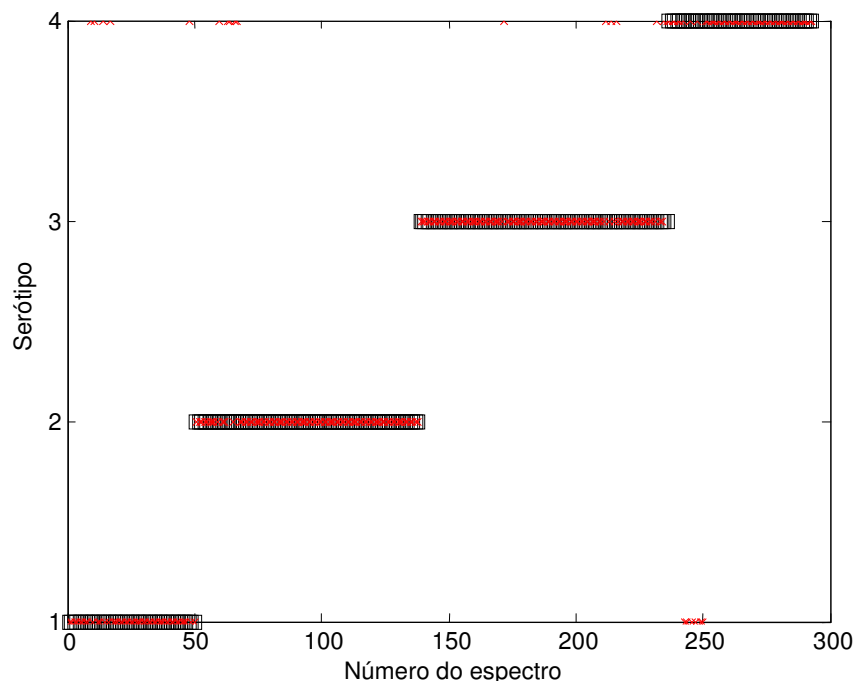


Figura III.36 - Modelo PLS hierárquico com 70% da totalidade dos espectros de FT-IR de *S. pneumoniae* dos quais foram eliminados os espectros aberrantes, os espectros da estirpe 2003V1484\$su19A e os espectros cujo serótipo não foi determinado pelo “método de xadrez”. Os espectros usados foram submetidos aos pré-processamentos: Savitzky-Golay (9,2,2) e SNV e usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número um corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”. A taxa de previsões correctas é de 92,2%.

Neste modelo, a taxa de identificação dos serótipos que coincide com a obtida pelo “método de xadrez” é de 92,2%.

A Figura III.37 mostra o resultado da validação do modelo PLS hierárquico para os 30% dos espectros não usados na sua construção.

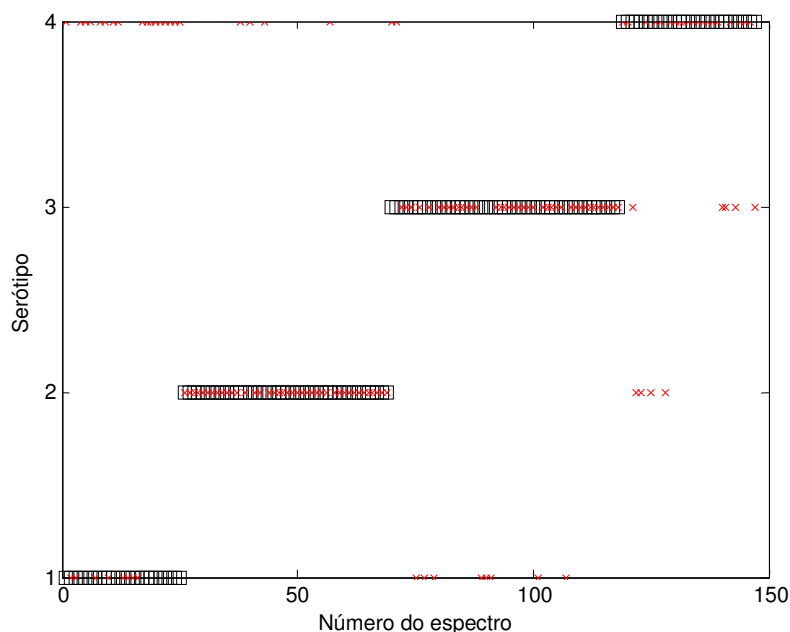


Figura III.37 - Validação do modelo PLS hierárquico com 30% da totalidade dos espectros de FT-IR de *S. pneumoniae* dos quais foram eliminados os espectros aberrantes e os espectros cujo serótipo não foi determinado pelo método clássico. Os espectros usados foram submetidos aos pré-processamentos: Savitzky-Golay (9,2,2) e SNV e usou-se a gama de números de onda de 900 a 1185 cm^{-1} . Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número um corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”. A taxa de previsões correctas é de 72,8%.

O modelo PLS hierárquico construído apresentou bons valores de sensibilidade e de especificidade e conseguiu atribuir correctamente o serótipo a 92,2% dos espectros usados na construção do modelo, no entanto, usando dados independentes para a validação, a percentagem de espectros com o serótipo correctamente atribuído foi de 72,8%, o que é um valor muito baixo. Assim, para a determinação do serótipo de uma estirpe de *S. pneumoniae* desconhecida, não deverá ser usado o modelo PLS hierárquico.

5.1.4. Redes neuronais

Para a optimização do método das redes neuronais, é necessária a definição dos pré-processamentos adequados de modo a que, no modelo

construído, as classes (serótipos) previstas correspondam às classes definidas previamente pelo “método de xadrez”. Existem outros parâmetros que têm de ser definidos tais como o número de nodos e o número de componentes principais a usar como entrada para a rede neuronal. Como o número de números de onda adquiridos é muito grande, torna-se necessário aplicar uma redução de dimensão aos nodos para ser usada a rede neuronal. Numa primeira abordagem foram usados todos os números de onda de todos os espectros definidos nas secções anteriores e como pré-processamentos usou-se o filtro de Savitzky-Golay (9,2,2) e SNV. Foram testadas várias combinações de número de nodos e de número de componentes principais. Os melhores resultados foram obtidos para um modelo com 6 nodos e 6 componentes principais (Figura III.38).

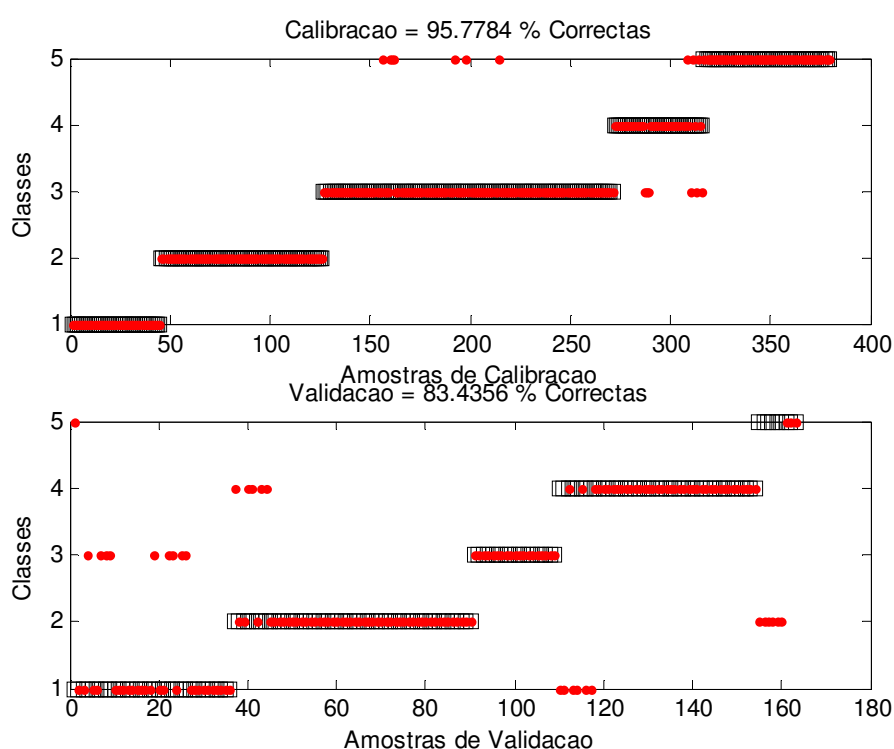


Figura III.38 – Representação gráfica dos resultados do modelo da rede neuronal aplicada à totalidade dos espectros. Foram usados como pré-processamentos o filtro de Savitzky-Golay (9,2,2) e SNV. A gama espectral usada foi de 900 a 1185 cm^{-1} . Representada com o símbolo (●) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”.

Tal como para as secções anteriores foi feito um estudo de redes neuronais para o conjunto de espectros resultantes da eliminação dos espectros aberrantes e dos espectros da estirpe 2003V1484\$su19A, mantendo o número de nodos de 6 e o número de componentes principais de 6. O resultado encontra-se na Figura III.39:

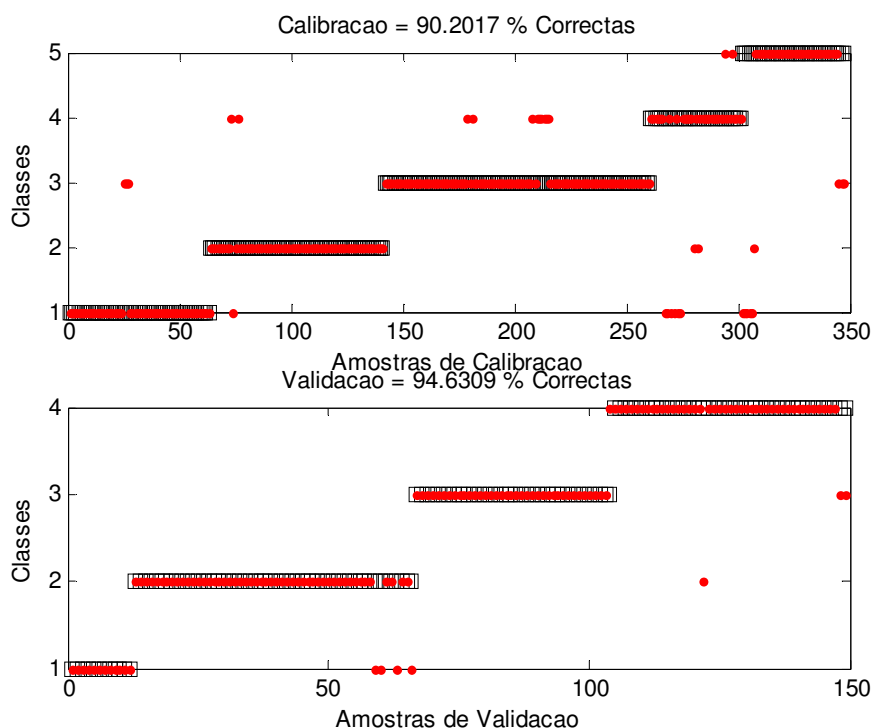


Figura III.39 - Representação gráfica dos resultados do modelo da rede neuronal aplicada aos espectros resultantes da eliminação dos espectros aberrantes e da estirpe 2003V1984\$su19A. Foram usados como pré-processamentos o filtro de Savitzky-Golay (9,2,2) e SNV. A gama espectral usada foi de 900 a 1185 cm^{-1} . Representada com o símbolo (\square) está o serótipo previsto pelo modelo e com o símbolo (\bullet) o serótipo definido pelo método clássico.

A validação está associada a uma baixa taxa de estirpes bem identificadas e, tal como para os modelos SIMCA e PLS, a estirpes pertencentes ao serótipo que não foi identificado pelo teste de referência, prejudicam a taxa de sucesso dos modelos. Assim, se se eliminarem estes espectros, o modelo que se obtém é o que se encontra na Figura III.40.

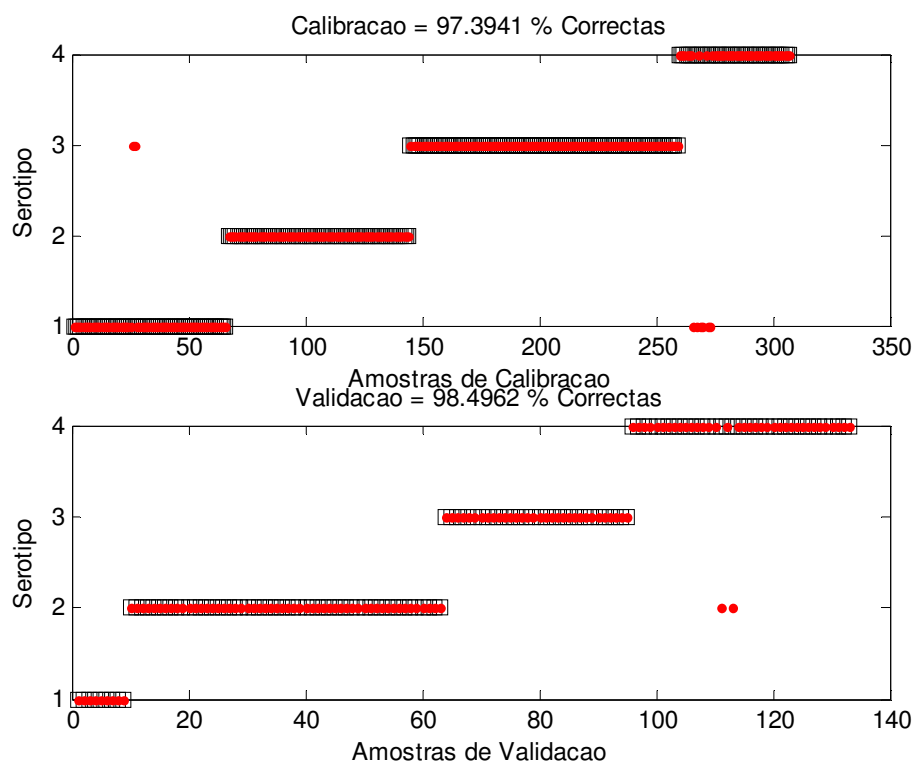


Figura III.40 - Representação gráfica dos resultados do modelo da rede neuronal aplicada aos espectros resultantes da eliminação dos espectros aberrantes e da eliminação dos espectros cujo serótipo não foi determinado pelo método clássico de serotipagem. Foram usados como pré-processamentos o filtro de Savitzky-Golay (9,2,2) e SNV. A gama espectral usada foi de 900 a 1185 cm^{-1} . Representada com o símbolo (●) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”.

Nas redes neuronais usadas, por cada aplicação da função, é gerado um modelo que é diferente do anterior, uma vez que a inicialização dos parâmetros da rede é feita de forma aleatória. Assim, de modo a estimar as taxas de identificação mais próximas da realidade, aquando da construção do modelo, e as taxas de atribuição correctas da validação do modelo definido, deve-se correr a função bastantes vezes e calcular a média dos 20 melhores modelos. Neste trabalho, foram construídos 60 modelos. Os resultados que se obtiveram estão na Tabela III.7.

Tabela III.7 – Resultados obtidos para um modelo de redes neuronais construído a partir de espectros de FT-IR de estirpes de *S. pneumoniae* que resultam da eliminação dos espectros aberrantes e dos espectros cujo serótipo não foi identificado pelo “método de xadrez”. Foram definidos 6 nodos e 6 componentes principais. Os valores representam a percentagem de classificação correcta para a construção do modelo e para a validação do mesmo. A sublinhado estão os modelos considerados para a média.

Modelo	Validação	Modelo	Validação	Modelo	Validação
99,0	84,2	96,4	82,7	96,1	88,7
98,0	81,2	97,7	78,2	98,0	96,2
99,7	86,5	97,4	92,5	97,1	89,5
100	82,7	96,7	80,5	97,4	85,0
98,0	88,7	97,7	85,0	96,4	84,2
100	84,2	100	85	97,1	94,7
96,4	88,0	97,4	88,7	99,3	90,2
97,4	98,5	97,4	74,4	97,4	92,5
96,4	94,0	100	85	98,4	85
98,4	85,0	97,7	91,7	99,0	85
96,1	90,2	98,0	86,5	96,4	92,5
100	86,5	97,7	85,0	100	87,2
97,1	92,5	98,0	89,5	98,0	79
98,0	83,5	97,7	77,4	97,1	92,5
97,7	87,2	98,0	77,4	100	89,5
99,7	88,7	97,7	96,2	99,7	85
96,4	77,4	95,8	85,7	98,4	86,5
98,4	81,2	98,0	85,0	99,0	86,5
99,3	89,5	100	86,5	97,7	82
95,8	92,5	98,0	71,4	96,1	88,7

Tendo em conta os modelos sublinhados da Tabela III.7, obtém-se como valores para a percentagem de identificação correcta da construção do modelo de 97,6 e para a validação uma percentagem de estirpes bem identificadas de 92,1. Tendo em conta estas percentagens, esta será também uma abordagem possível para a serotipagem de estirpes pertencentes a um dos 4 serótipos definidos.

5.2 GC

As cadeias polissacáridas da parede celular encontram-se ligadas a ácidos gordos e, o objectivo de fazer um estudo de ácidos gordos nas estirpes de *S. pneumoniae*, é tentar perceber se existe uma relação entre a composição

em ácidos gordos e os serótipos. Mais uma vez, os serótipos estudados foram quatro e um quinto grupo de estirpes onde o serótipo não foi determinado com sucesso pelo “método de xadrez” (estirpes descritas na Tabela II.1).

Os resultados resultam de dois dias de extracção dos ácidos gordos das estirpes de *S. pneumoniae* e de uma análise no cromatógrafo gasoso também de dois dias distintos. Um exemplo do relatório que se obtém da análise dos ácidos gordos usando o método Sherlock[®], encontra-se na Figura III.41.

No anexo III encontra-se uma tabela (Tabela VII.1) com os resultados da identificação e percentagem dos ácidos gordos encontrados e o tempo de retenção para cada ácido gordo, para as estirpes analisadas.

Como os resultados do GC se traduzem em presença ou ausência de um dado ácido gordo, ou seja, são dados discretos, não se devem aplicar outros pré-processamentos que não sejam o da remoção da média ou o da autonormalização, que pretendem eliminar diferenças que advêm de concentrações diferentes de microrganismos na suspensão inicial. O efeito da concentração deve ser eliminado pois, consoante a resposta total de um dado microrganismo, assim podem ser visíveis ácidos gordos que estão presentes em quantidades muito pequenas, mas que não dependem do serótipo.

Fazendo um estudo de componentes principais para as estirpes estudadas, submetidas ao pré-processamento de remoção da média ou autonormalização e, fazendo o cruzamento com os serótipos o resultado a que se chega é semelhante para ambos os pré-processamentos. Na Figura III.42 encontra-se a representação gráfica do PCA de resultados submetidos à remoção da média.

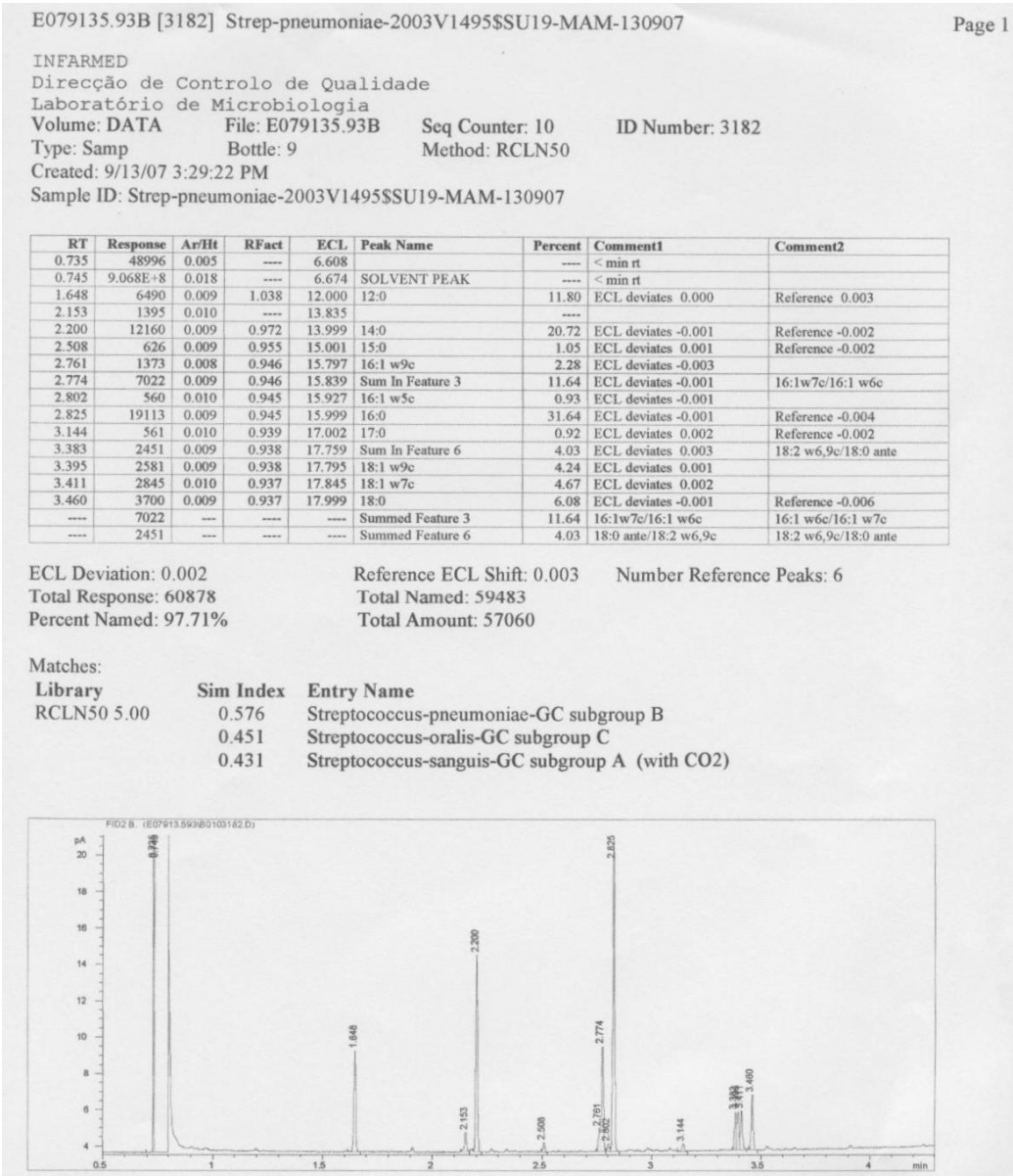


Figura III.41– Relatório que se obtém após a análise de ácidos gordos por GC de uma amostra da estirpe 2003V1495\$su19A, pelo software Sherlock®.

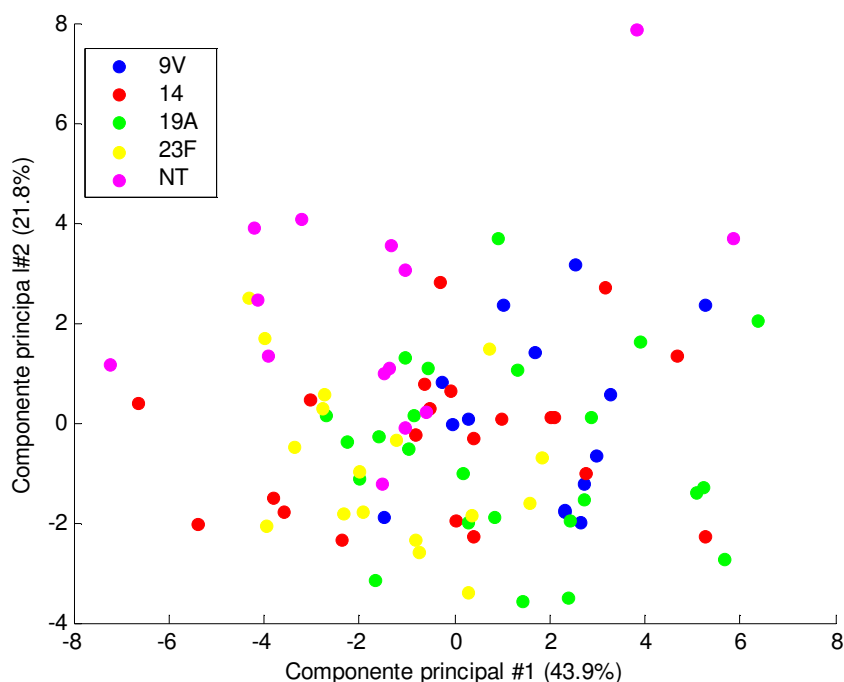


Figura III.42 – Análise de componentes principais para estirpes de *S. pneumoniae* cujos ácidos gordos foram extraídos e submetidos a uma análise por GC. Para esta análise os resultados foram submetidos a um pré-processamento de remoção da média.

Observando a Tabela VII.1 (anexo III) verifica-se que existem colunas onde praticamente não existe expressão de ácidos gordos e, nos casos em que existe, a quantidade é muito pequena, o que acontece para as estirpes cuja concentração da suspensão inicial foi elevada. Assim, estes valores estão relacionados com a concentração da suspensão inicial e não com uma característica intrínseca da estirpe logo, as colunas correspondentes aos ácidos gordos 13:0, 13:0 2OH, 15:1 w5c, 16:1 w5c, 17:1 iso w5c, 17:1 anteiso w9c, 17:1 w8c, 17:0, 18:1 w5c, 19:1 iso, 19:0 iso e 20:4 w6,9,12,15c, foram eliminadas. Foram também retirados os espectros considerados aberrantes pela estatística de Hotelling e pela estatística dos resíduos e os resultados respeitantes às estirpes cujo serótipo não foi determinado pelo método de referência. Construindo um novo gráfico que traduz a análise de PCA para os dados resultantes, o que se obtém é o representado na Figura III.43.

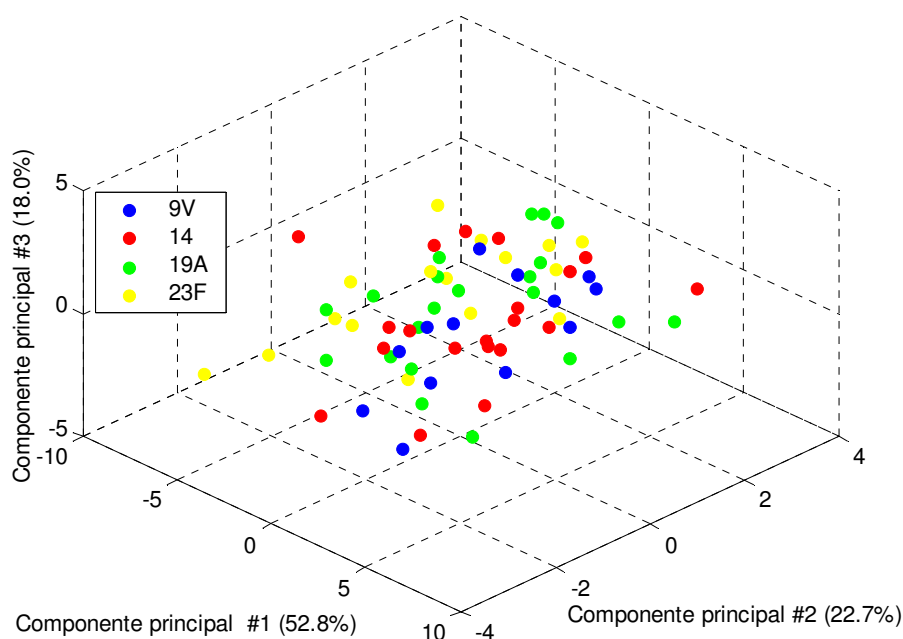


Figura III.43 - Análise de componentes principais a 3 dimensões para estirpes de *S. pneumoniae* cujos ácidos gordos foram extraídos e submetidos a uma análise por GC. Para esta análise os resultados foram submetidos a um pré-processamento de remoção da média e foram eliminados os resultados referentes às amostras sem serótipo determinado pelo teste de referência para a serotipagem.

Como se pode observar na figura, não existe um agrupamento claro das estirpes de acordo com os seus serótipos.

Partindo do mesmo conjunto de resultados de GC resultantes da eliminação dos espectros aberrantes e dos espectros sem serótipo determinado pelo método de referência, estudou-se um modelo SIMCA usando como pré-processamentos os usados na análise de PCA. Os resultados obtidos para todos os pré-processamentos foram semelhantes. Na Figura III.44 apresenta-se a representação gráfica para o pré-processamento com remoção da média e para 2 componentes principais que, de acordo com os resultados obtidos para a validação cruzada para 5 blocos contíguos, é o que está associado a um erro menor.

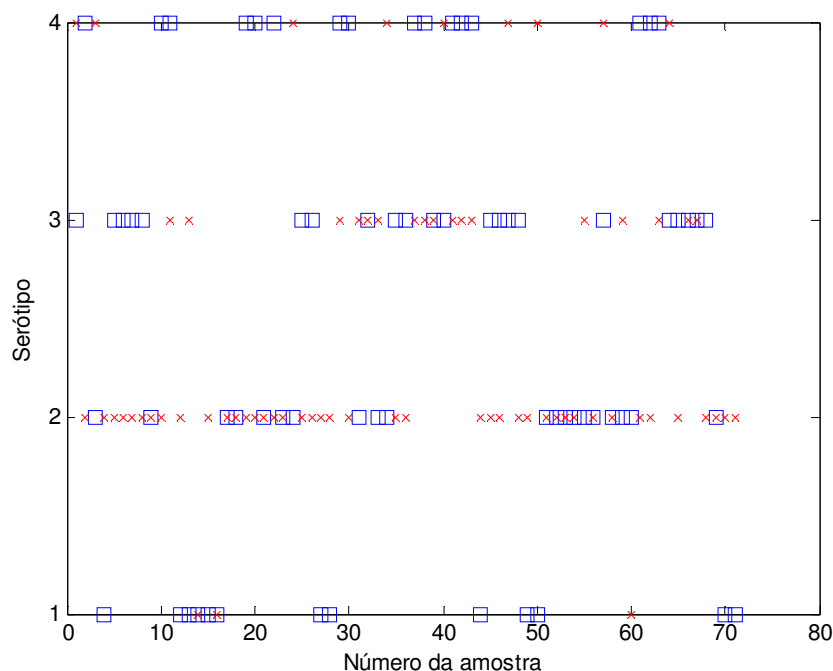


Figura III.44 – Resultado da validação cruzada para 5 blocos contíguos para o modelo SIMCA para estirpes de *S. pneumoniae* cujos ácidos gordos foram extraídos e submetidos a uma análise por GC. Para a construção do modelo os resultados foram submetidos a um pré-processamento de remoção da média e foram usados 3 componentes principais. Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número um corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A, o 4 ao 23F e o 5 ao serótipo não determinado. Representado com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo método de referência.

O resultado do modelo SIMCA para a serotipagem indica que não foi possível identificar com sucesso o serótipo para 74,7% dos resultados, o que não é satisfatório. Se se dividirem os resultados de GC de modo a que 70% dos resultados seja usado para a construção de um modelo SIMCA e 30% para a sua validação, o que se obtém como resultado da validação é o representado na Figura III.45.

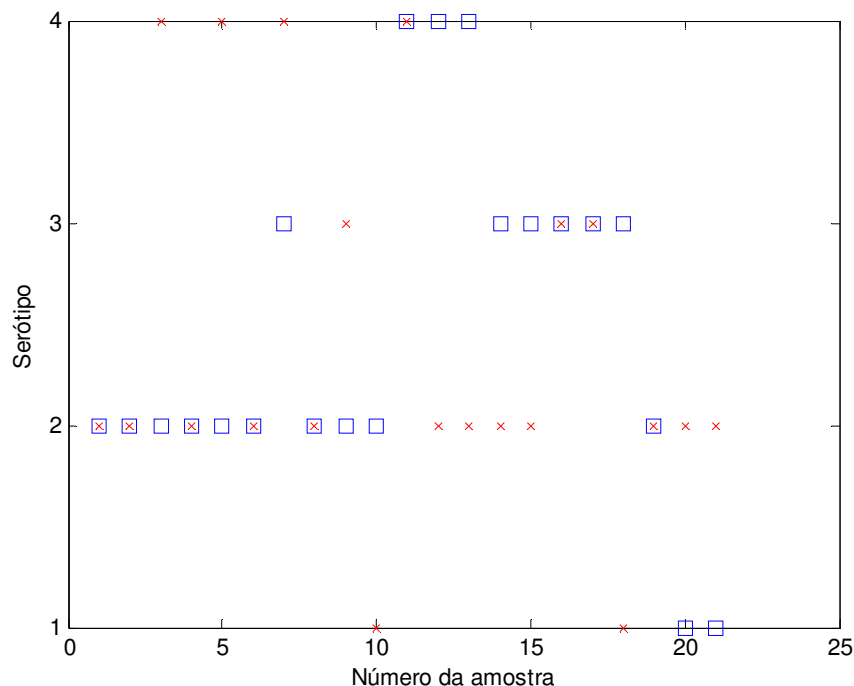


Figura III.45– Resultado da validação do modelo SIMCA para 30% dos resultados da análise de GC para estirpes de *S. pneumoniae*. Para a construção do modelo foram usados os resultados que foram submetidos a um pré-processamento de remoção da média e foram usados 2 componentes principais. Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número um corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representado com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo método clássico. O modelo apenas conseguiu prever correctamente o serótipo de 9 de 21 resultados (42,9%).

Neste caso o modelo conseguiu prever o serótipo concordante com o do “método de xadrez” para apenas 42,9% dos resultados.

Um outro modelo supervisionado a testar é o PLS discriminante. Para a elaboração deste modelo partiram-se dos resultados obtidos pela análise de ácidos gordos por GC, que resultam da eliminação dos espectros aberrantes e dos espectros sem serótipo determinado pelo método de referência e que foram submetidos aos pré-processamentos de remoção da média ou autonormalização. O modelo obtido (Figura III.46) resulta de uma validação cruzada com 8 blocos contíguos e 5 variáveis latentes.

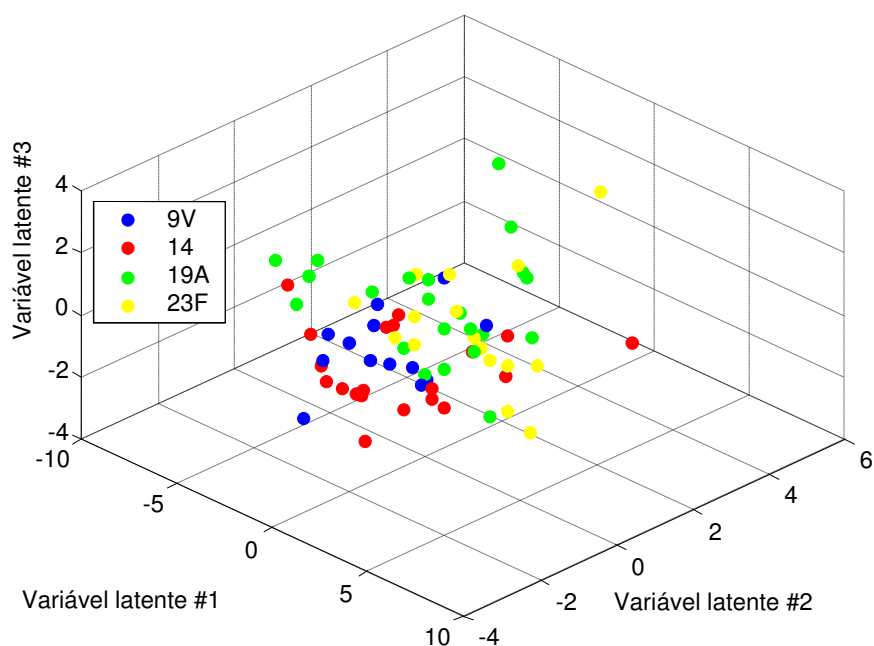


Figura III.46 - Resultados obtidos para o modelo de PLS discriminante para o conjunto de resultados da análise de ácidos gordos por GC de *S. pneumoniae*, dos quais foram eliminados os resultados aberrantes e os resultados das estirpes cujo serótipo não foi determinado pelo método de referência. Este modelo resulta de uma validação cruzada com 8 blocos contíguos e 5 variáveis latentes. Os resultados de GC usados foram submetidos a uma remoção da média e foram usados os ácidos gordos com quantidades significativas. A equivalência entre as cores e os serótipos encontra-se na legenda.

A representação gráfica a 3 dimensões da Figura III.46 mostra que não existe uma separação nítida dos serótipos. Mas, de modo a avaliar o modelo, os dados foram divididos em dois grupos um para construção do modelo com 50 resultados de GC (70%) e outro para validação do modelo com 21 resultados de GC (30%). A Figura III.47 mostra o resultado obtido para a validação com 30% das estirpes.

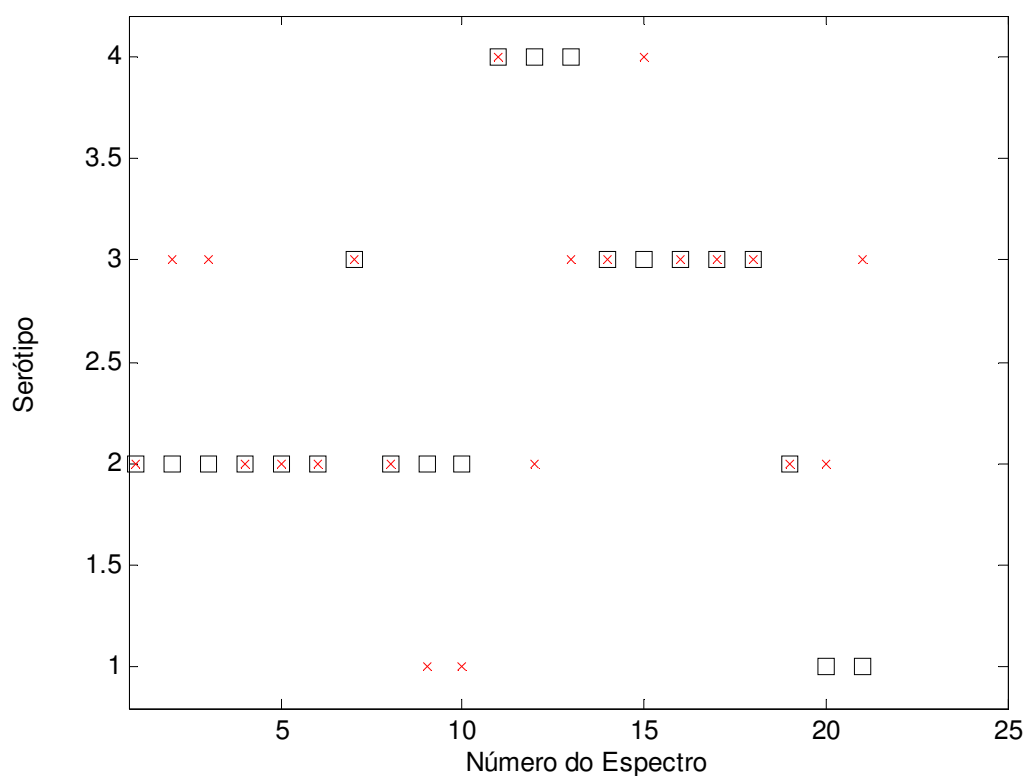


Figura III.47 - Resultados obtidos para a validação do modelo de PLS discriminante para 30% do conjunto de resultados da análise de ácidos gordos por GC de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes e os espectros cujo serótipo não está determinado. Este modelo resulta de uma validação cruzada com 5 variáveis latentes. Os resultados foram submetidos à remoção da média e usaram-se apenas as colunas com quantidades de ácidos gordos significativas. Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo método de referência. Com a designação 1 está o serótipo 9V, 2 o serótipo 14, 3 o serótipo 19A e 4 o serótipo 23F. A percentagem de estirpes correctamente identificadas foi de 57,1.

O modelo PLS discriminante conseguiu apenas identificar correctamente 57,1% dos espectros usados na validação.

Outro modelo supervisionado é o PLS hierárquico. Para o PLS hierárquico, foram testados os pré-processamentos referidos para os modelos SIMCA e PLS discriminante; escolheram-se 5 blocos contíguos para a validação cruzada e usaram-se os resultados resultantes da análise de ácidos gordos por GC, aos quais foram eliminados os resultados aberrantes e os resultados das estirpes cujo serótipo não foi determinado pelo método de

referência. O melhor resultado obtido foi para o pré-processamento da remoção da média e o resultado encontra-se na Tabela III.8:

Tabela III.8 - Resultados obtidos para o modelo de PLS hierárquico, para 5 blocos contíguos, para os resultados da análise dos ácidos gordos por GC de *S. pneumoniae*. Os resultados usados foram submetidos à remoção da média. Nesta tabela encontram-se representados os serótipos, o número de variáveis latentes usadas para a identificação de cada serótipo, a sensibilidade e a especificidade.

Serótipo	LV	Sensibilidade (α)	Especificidade (β)
9V	5	81,8	94,9
23F	4	76,9	88,5
14	4	80,0	50,0
19A	-	100,0	100,0

Os valores de sensibilidade e especificidade para o modelo são baixos. Quando se dividem os resultados em 70% para construção do modelo de PLS hierárquico e 30% para validação do modelo obtido, o resultado para a validação de percentagem de estirpes com serótipo correctamente atribuído é de 47,6, como se vê na Figura III.48. Este resultado não se considera satisfatório.

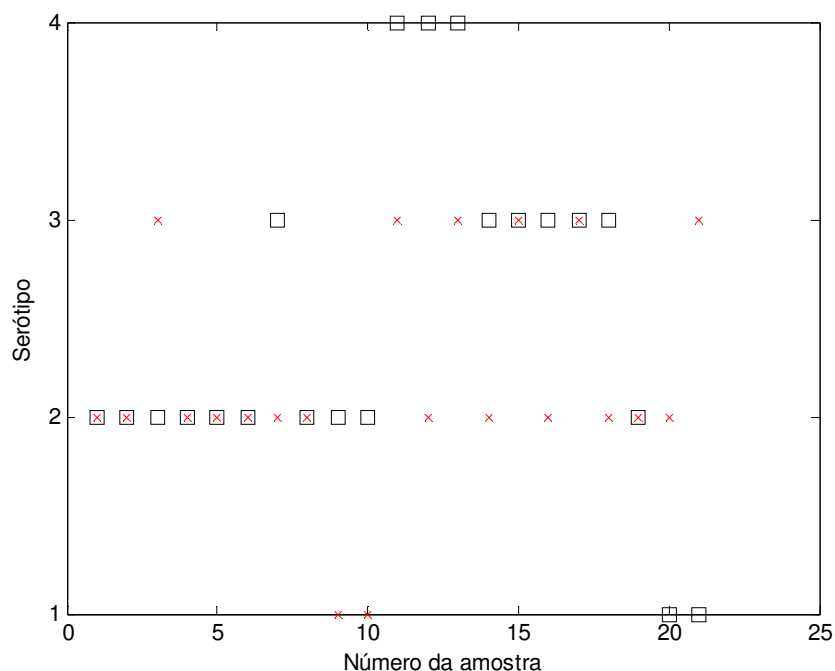


Figura III.48 - Validação do modelo PLS hierárquico com 30% do conjunto de resultados da análise de ácidos gordos por GC de *S. pneumoniae* dos quais foram eliminados os resultados aberrantes e os resultados das estirpes cujo serótipo não foi determinado pelo método de referência. Os resultados de GC foram submetidos a uma remoção da média e foram eliminados os ácidos gordos cuja expressão não é significativa. Neste gráfico encontram-se representados no eixo das ordenadas os serótipos do *S. pneumoniae*, onde o número um corresponde ao serótipo 9v, o 2 ao 14, o 3 ao 19A e o 4 ao 23F. Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo teste de referência. A taxa de previsões correctas é de 47,6%

Quando se aplicam as redes neuronais ao conjunto total de dados descrito nos modelos anteriores, o que se obtém é o representado na Figura III.49.

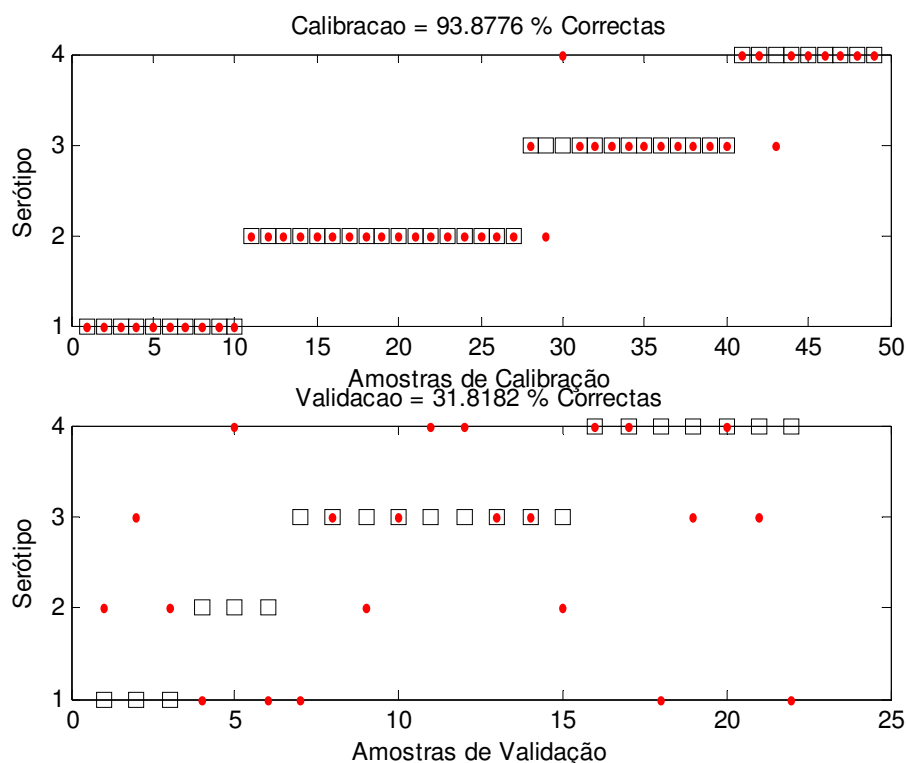


Figura III.49 - Representação gráfica dos resultados do modelo da rede neuronal aplicada aos resultados da análise de ácidos gordos por GC de *S. pneumoniae*. Foi usado como pré-processamento a remoção de média. Foram usados todos os ácidos gordos, 6 nodos e 6 componentes principais. Representada com o símbolo (•) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”. No eixo das ordenadas estão os serótipos cuja equivalência é a seguinte: 1 corresponde ao serótipo 9V, o 2 ao serótipo 14, o 3 ao serótipo 19A e o 4 ao serótipo 23F.

Para este modelo a percentagem de estirpes com o serótipo correctamente atribuído para as estirpes de validação foi de apenas 31,8%, o que é um resultado bastante baixo.

Tendo em conta os resultados obtidos para a validação dos modelos construídos, considera-se que não existe uma relação entre os ácidos gordos de cadeia C9 a C20 e os serótipos de *S. pneumoniae*, detectável pelos modelos testados e nas condições experimentais descritas.

6. Susceptibilidade à Penicilina

6.1. FT-IR

6.1.1 PCA

Tal como para o estudo da serotipagem dos pneumococos optou-se por iniciar o estudo da susceptibilidade à penicilina por uma análise de componentes principais. Quando se usa todo o espectro para a análise, não se recorrendo a pré-processamentos, e se usa um número de componentes principais de 9 (valor determinado pela representação dos valores próprios), o que se obtém é o gráfico representado na Figura III.50:

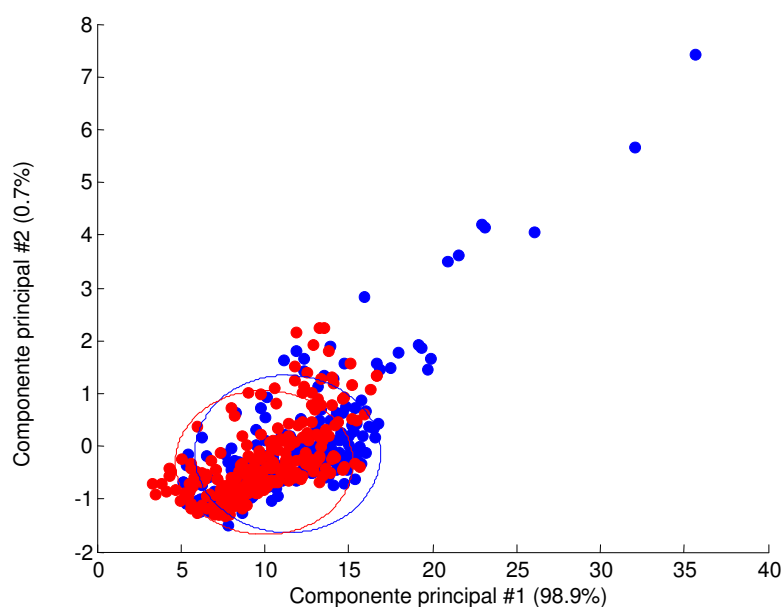


Figura III.50 - Análise de componentes principais para os 542 espectros, tal como foram obtidos pela análise de FT-IR, para estirpes de *S. pneumoniae*. Espectros das estirpes sensíveis (●) e espectros das estirpes resistentes à penicilina (●).

Numa segunda abordagem recorreu-se a todos os números de onda dos espectros, a um número de componentes principais de 4 e vários pré-processamentos (filtro de Savitzky-Golay (17,2,1) + remoção da média,

filtro de Savitzky-Golay (17,2,2) + remoção da média, SNV + remoção da média, filtro de Savitzky-Golay (9,2,2) + SNV e, MSN). Como se comprova pela observação da Figura III.51, não foi possível a distinção das estirpes de acordo com a sua susceptibilidade à penicilina para nenhuma das condições referidas, recorrendo às várias combinações de componentes principais.

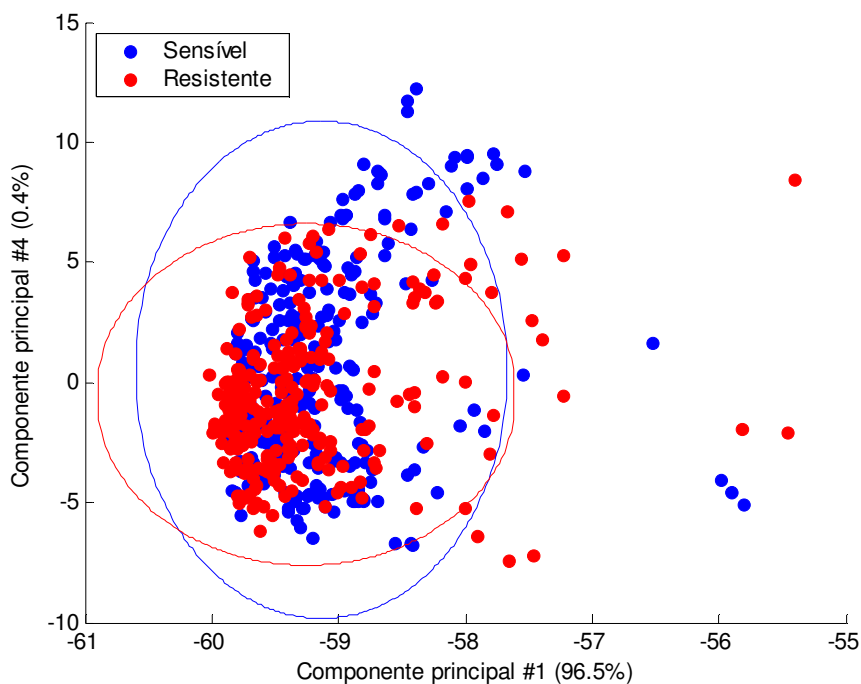


Figura III.51 - Gráfico de componentes principais para 4 componentes principais e para todos os espectros analisados, utilizando todos os números de onda, com um pré-processamento de Savitzky-Golay (9,2,2) e SNV. Espectros das estirpes sensíveis (●) e espectros das estirpes resistentes à penicilina (●).

A resistência dos *S. pneumoniae* deve-se à alteração das enzimas PBP que conseguem sintetizar a parede celular, mesmo na presença de antibiótico. Assim, foram estudadas as zonas espectrais correspondentes à parede celular (números de onda de 900 a 1185 cm^{-1}), à amida I, à amida II e à amida III e às 3 zonas espectrais das 3 amidas em simultâneo. Para todas as zonas espectrais foram testados os pré-processamentos: filtro de Savitzky-Golay (17,2,1) + remoção da média, filtro de Savitzky-Golay (17,2,2) +

remoção da média, SNV + remoção da média e filtro de Savitzky-Golay (9,2,2) + SNV. Os resultados obtidos não foram satisfatórios, como se pode ver na Figura III.52, que é representativa de todos os resultados obtidos para os vários pré-processamentos, para as várias zonas espectrais.

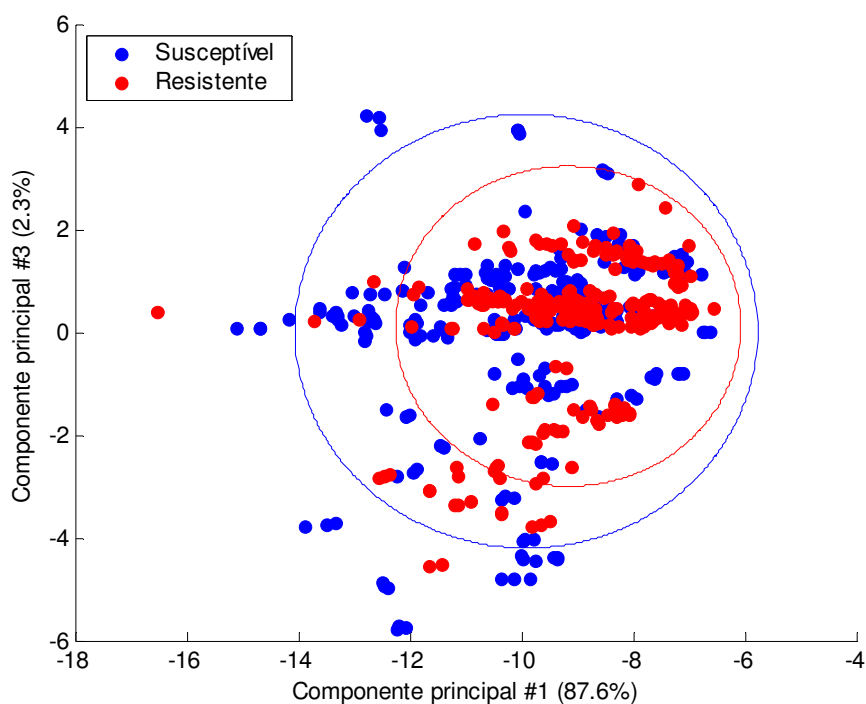


Figura III.52- Gráfico de componentes principais para todos os espectros analisados, utilizando a zona da gama espectral de 900 a 1185 cm^{-1} , com um pré-processamento de Savitzky-Golay (9,2,2) e SNV. Espectros das estirpes sensíveis (•) e espectros das estirpes resistentes à penicilina (•).

6.1.2. SIMCA

Tal como para o PCA, não foi possível construir um modelo SIMCA para avaliar a susceptibilidade das estirpes de *S. pneumoniae* analisadas. Este facto seria de esperar pois a construção do modelo SIMCA pressupõe um estudo prévio dos pré-processamentos adequados que é feito na análise de PCA. Como nesta última análise não se conseguiu determinar as condições adequadas à correcta separação das estirpes em sensível e resistente à

penicilina, o mesmo aconteceu aquando da elaboração o modelo SIMCA. Um exemplo do que se obteve é o que se encontra na Figura III.53.

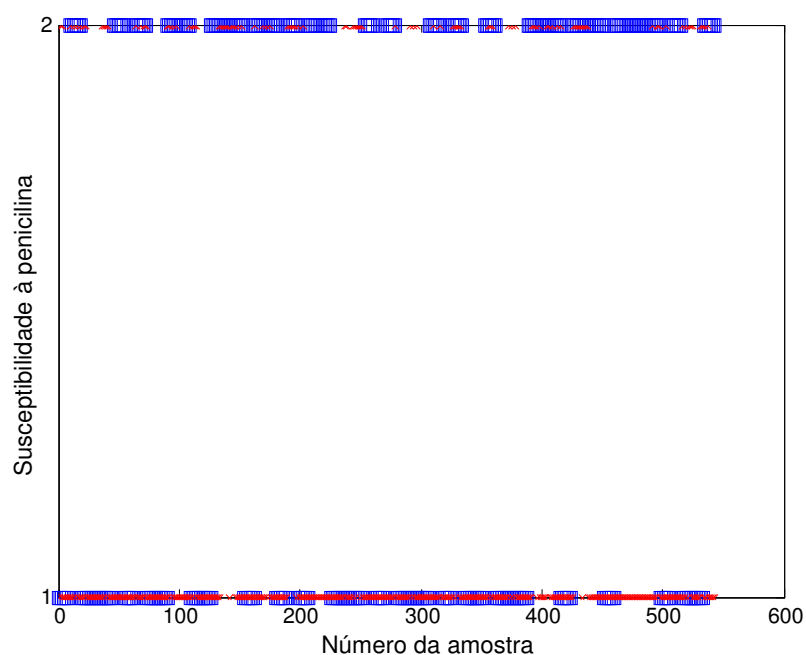


Figura III.53 – Resultado da validação cruzada para o modelo SIMCA com 14 componentes principais para o conjunto total de espectros. Neste gráfico no eixo das ordenadas está a susceptibilidade das estirpes de *S. pneumoniae*, onde o número 1 corresponde às estirpes sensíveis e 2 às estirpes resistentes. Com o símbolo (x) encontra-se a classe prevista pelo modelo e com o símbolo (□) a classe definida previamente, para cada resultado. Existe uma percentagem de espectros mal classificadas de 42,6%.

Neste exemplo existe uma taxa de estirpes mal classificadas de 42,6%, o que não se considera aceitável para a construção de um modelo.

Se se considerarem gamas de números de onda, tal como foi feito para a análise de PCA, e os vários pré-processamentos referidos, o resultado a que se chega é semelhante para todas as zonas espectrais: amida I, amida II, amida III, parede celular e amida I+amida II+ amida III. Na Figura III.54 mostra-se o resultado obtido para a zona espectral da amida I.

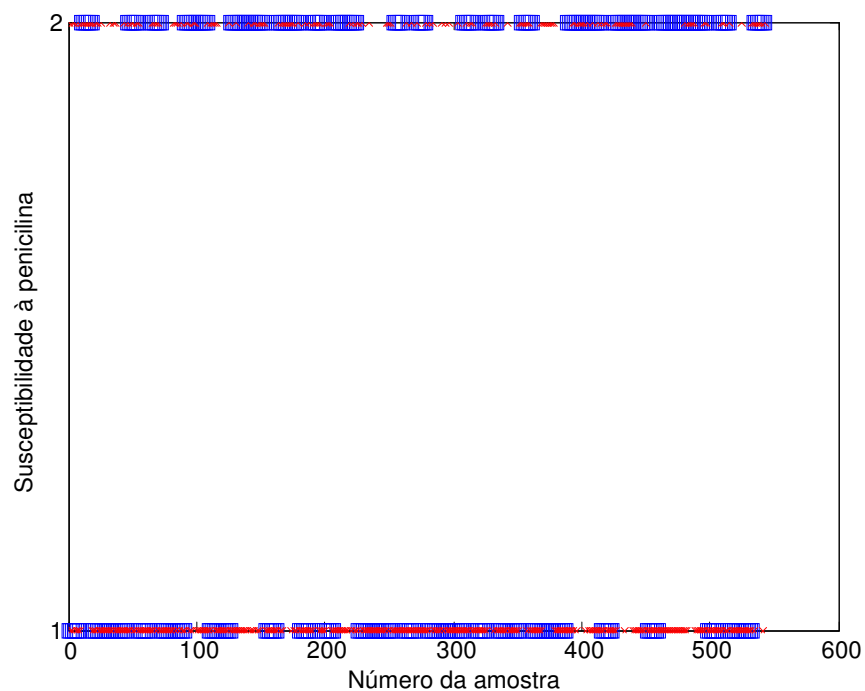


Figura III.54 - Resultado da validação cruzada para o modelo SIMCA com 19 componentes principais para o conjunto total de espectros e para a gama espectral da Amida I. Neste gráfico no eixo das ordenadas está a susceptibilidade das estirpes de *S. pneumoniae*, onde o número 1 corresponde às estirpes sensíveis e 2 às estirpes resistentes. Com o símbolo (x) encontra-se a classe prevista pelo modelo e com o símbolo (□) a classe definida previamente, para cada resultado. Existe uma percentagem de espectros mal classificadas de 40,0%.

A percentagem de estirpes mal classificadas foi de 40,0%. Assim, pode-se concluir que não é possível definir um modelo SIMCA que avalie a susceptibilidade à penicilina, considerando apenas uma gama espectral e os pré-processamentos de SNV e filtro de Savitzky-Golay com os vários parâmetros do filtro atrás descritos.

6.1.3. PLS

6.1.3.1 PLS discriminante

Para o método supervisionado PLS discriminante, tal como para o PCA e SIMCA, é necessária a definição do pré-processamento adequado à correcta separação dos espectros consoante a sua susceptibilidade à penicilina. Assim, foram testados modelos com os pré-processamentos: filtro de

Savitzky-Golay (17,2,1), filtro de Savitzky-Golay (17,2,2), filtro de Savitzky-Golay (9,2,1), filtro de Savitzky-Golay (9,2,2), SNV, filtro de Savitzky-Golay (17,2,1) + SNV, filtro de Savitzky-Golay (17,2,2) + SNV, filtro de Savitzky-Golay (9,2,1) + SNV e filtro de Savitzky-Golay (9,2,2) + SNV. Foi usada para a construção do modelo a totalidade dos espectros e toda a gama espectral. Para se determinar o número de variáveis latentes adequada recorreu-se à validação cruzada com 10 blocos contíguos. Os resultados obtidos foram muito semelhantes apresentando-se, na Figura III.55, o resultado para o pré-processamento filtro de Savitzky-Golay (9,2,2) e SNV.

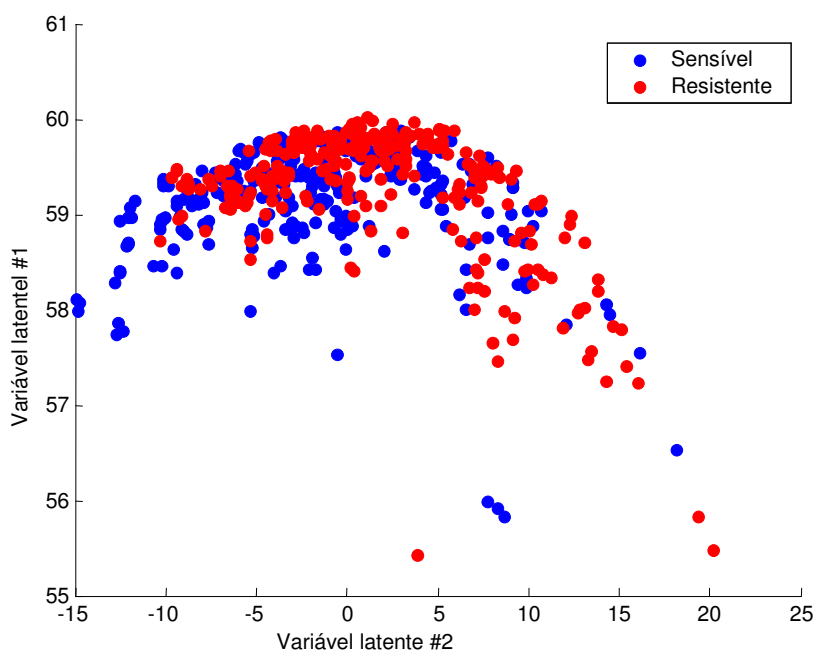


Figura III.55 - Resultados obtidos para o modelo PLS discriminante usando para a validação cruzada 10 blocos contíguos e 6 variáveis latentes. Foram usados todos os espectros FT-IR de *S. pneumoniae*, que foram submetidos a pré-processamentos filtro de Savitzky-Golay (9,2,2) e SNV. A equivalência entre as cores e a susceptibilidade encontra-se na legenda.

De modo a tentar otimizar os resultados, foram testadas várias gamas espectrais (amida I, amida II, amida III, conjunto das zonas espectrais da amida I+amida II+amida III e zona da parede celular (900 a 1185 cm^{-1}))

combinadas com os vários pré-processamentos referidos no parágrafo anterior. O resultado obtido foi, mais uma vez, semelhante para todas as gamas espectrais. Na Figura III.56 encontra-se a representação gráfica a três dimensões para a zona espectral da amida I.

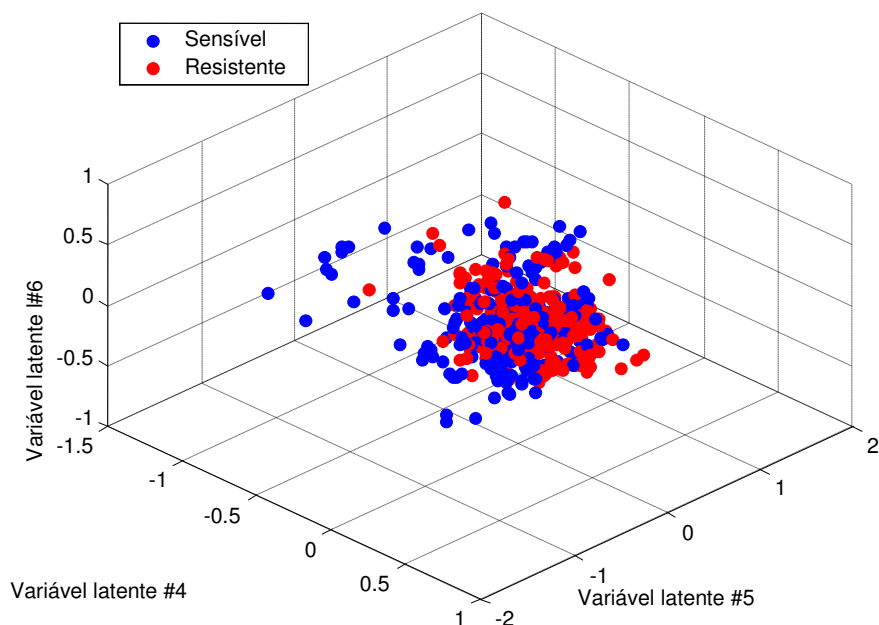


Figura III.56 - Resultados obtidos para o modelo de PLS discriminante usando para a validação cruzada 10 blocos contíguos e 6 variáveis latentes. Foram usados todos os espectros FT-IR de *S. pneumoniae*, que foram submetidos a pré-processamentos filtro de Savitzky-Golay (9,2,2) e SNV. Foi usada a gama espectral correspondente à amida I. A equivalência entre as cores e a susceptibilidade encontra-se na legenda.

Tal como o representado na Figura III.56, nenhuma das zonas espectrais estudadas permitiu construir um modelo que separasse correctamente os espectros de acordo com a susceptibilidade à penicilina logo, este modelo não poderá ser usado com este fim.

6.1.3.2. PLS hierárquico

Uma vez que não foi possível construir um modelo com método supervisionado SIMCA e PLS discriminante, tentou-se construir um modelo com outro método supervisionado, o PLS hierárquico. Assim, construíram-se modelos para a totalidade dos espectros utilizando cada um dos seguintes pré-processamentos: filtro de Savitzky-Golay (17,2,1), filtro de Savitzky-Golay (17,2,2), filtro de Savitzky-Golay (9,2,1), filtro de Savitzky-Golay (9,2,2), SNV, filtro de Savitzky-Golay (17,2,1) + SNV, filtro de Savitzky-Golay (17,2,2) + SNV, filtro de Savitzky-Golay (9,2,1) + SNV e filtro de Savitzky-Golay (9,2,2) + SNV. Os melhores resultados foram para o pré-processamento Savitzky-Golay (9,2,2) + SNV. Apenas se apresenta o resultado do modelo PLS hierárquico com 16 blocos contíguos para a validação cruzada e com este último pré-processamento (Tabela III.9).

Tabela III.9 – Resultados obtidos para o modelo PLS hierárquico para todos os espectros FT-IR de *S. pneumoniae*. Os espectros usados foram submetidos a pré-processamentos Savitzky-Golay (9,2,2) e SNV e usou-se a totalidade dos números de onda. Nesta tabela estão duas linhas onde uma corresponde às estirpes resistentes e a outra às estirpes sensíveis, cada uma com o número de variáveis latentes usadas para a identificação da susceptibilidade, a sensibilidade e a especificidade do modelo.

Teste de Susceptibilidade	LV	Sensibilidade (α) (%)	Especificidade (β) (%)
Sensível	15	80,37	47,43
Resistente	0	100	100

Uma vez que o modelo construído tem uma sensibilidade e especificidade muito baixas testaram-se modelos com os pré-processamentos: referidos no parágrafo anterior apenas para as zonas espectrais correspondentes à amida I, à amida II, à amida III, ao conjunto das zonas espectrais da amida I+amida II+amida III e para a zona espectral da parede celular (900 a 1185 cm^{-1}). Os melhores resultados ocorreram para o modelo

com o pré-processamento Savitzky-Golay (9,2,2) + SNV. A sensibilidade e especificidade do modelo construído para a zona espectral da amida I foram semelhantes às obtidas para as restantes zonas espectrais estudadas. Assim, apenas se apresenta o resultado do modelo para a amida I e para 16 blocos contíguos.

Tabela III.10 – Resultados obtidos para o modelo PLS hierárquico para 16 blocos contíguos para todos os espectros FT-IR de *S. pneumoniae*. Os espectros usados foram submetidos a pré-processamentos Savitzky-Golay (9,2,2) e SNV e usou-se a zona espectral correspondente à da amida I. Nesta tabela estão duas linhas onde uma corresponde às estirpes resistentes e a outra às estirpes sensíveis cada uma com o número de variáveis latentes usadas para a identificação da susceptibilidade, a sensibilidade e a especificidade do modelo.

Teste de Susceptibilidade	LV	Sensibilidade (α) (%)	Especificidade (β) (%)
Sensível	3	50,37	70,96
Resistente	0	100	100

Dos resultados obtidos conclui-se que o modelo PLS hierárquico, aplicado a espectros de FT-IR nas condições estudadas neste trabalho, não permite avaliar a susceptibilidade à penicilina dos pneumococos.

6.1.4. Redes neuronais

Numa primeira abordagem foram usados todos os números de onda de todos os espectros e os pré-processamentos SNV, SNV com filtro de Savitzky-Golay com os seguintes parâmetros: (9,2,1), (9,2,2), (17,2,1) e (17,2,2) combinados com nodos de valores e componentes principais. O resultado obtido encontra-se na Figura III.57.

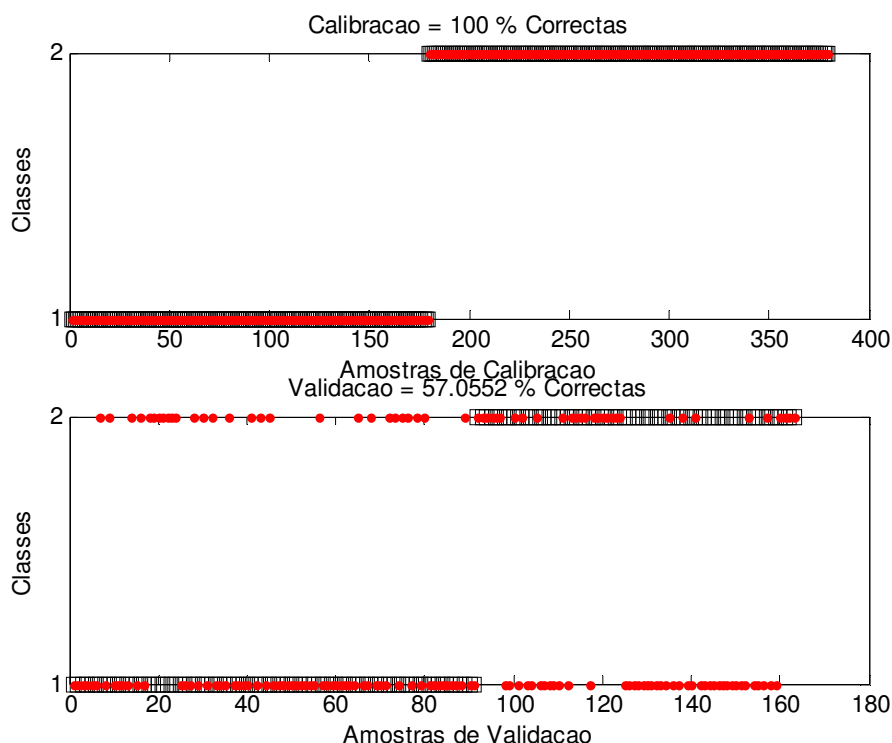


Figura III.57 - Representação gráfica dos resultados do modelo da rede neuronal aplicada a todos os espectros de FT-IR de *S. pneumoniae*. Foram usados como pré-processamentos o filtro de Savitzky-Golay (9,2,2) e SNV. Foi usada toda a gama espectral, 10 nodos e 10 componentes principais. Representada com o símbolo (•) está a classe resistente/sensível valor prevista pelo modelo e com o símbolo (◻) a classe resistente/sensível definida pelo método CLSI. A classe sensível corresponde ao número 1 e a resistente ao número 2.

Os dados de validação indicam que o modelo conseguiu atribuir a classe resistente ou sensível semelhante ao resultado da susceptibilidade à penicilina pelo método do CLSI para apenas 57,1% dos espectros. Seleccionando zonas espectrais: Amida I, Amida II, Amida III, Amida I+Amida II+Amida III e, zona da parede e, os pré-processamentos descritos atrás, o que se obtém é um resultado semelhante para todas as gamas espectrais. Assim, na Figura III.58 apresenta-se como exemplo o modelo das redes neuronais aplicada a todos os espectros com a gama espectral correspondente à zona da Amida I com 10 nodos e 10 componentes principais e um pré-processamento de filtro de Savitzky-Golay (9,2,2) e SNV.

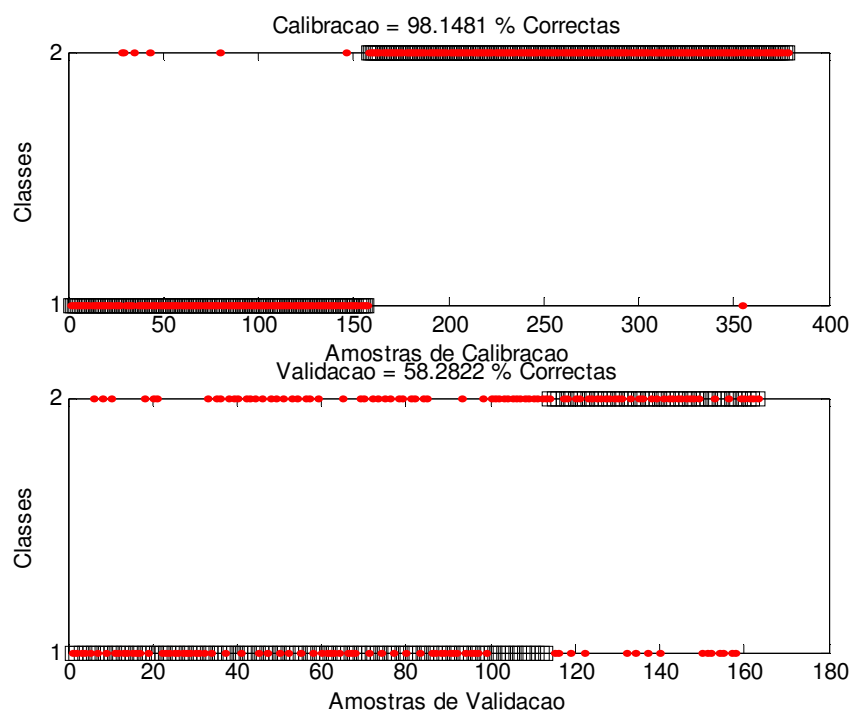


Figura III.58 - Representação gráfica dos resultados do modelo da rede neuronal aplicada a todos os espectros de FT-IR de *S. pneumoniae*. Foram usados como pré-processamentos o filtro de Savitzky-Golay (9,2,2) e SNV. Foi usada toda a gama espectral da Amida I, 10 nodos e 10 componentes principais. Representada com o símbolo (•) está a classe resistente/sensível valor prevista pelo modelo e com o símbolo (□) a classe resistente/sensível definida pelo método CLSI. A classe sensível corresponde ao número 1 e a resistente ao número 2.

Pode-se concluir que, também, com o modelo das redes neuronais, não é possível estudar a susceptibilidade à penicilina das estirpes de *S. pneumoniae*, nas condições estudadas nesta dissertação.

6.2 GC

Com esta análise pretende-se estudar a possível relação entre a susceptibilidade à penicilina e os ácidos gordos de cadeia C9 a C20 presentes na célula. Assim, numa primeira fase foi feita uma análise de componentes principais aos dados obtidos da extracção dos ácidos gordos constantes na Tabela VII.1 (anexo III). Como pré-processamentos testou-se a remoção da média e a autonormalização. Para nenhum dos pré-processamentos se

conseguiu uma separação dos resultados em dois grupos consoante a susceptibilidade, testando as várias combinações de componentes principais. Um exemplo do que se obteve para um modelo com 5 componentes principais e usando como pré-processamento a remoção da média encontra-se na Figura III.59. Neste exemplo está representado o componente principal número 3 *versus* o componente principal número 1 pois foi aquele que apresentou uma melhor separação entre classes, se bem que não satisfatória.

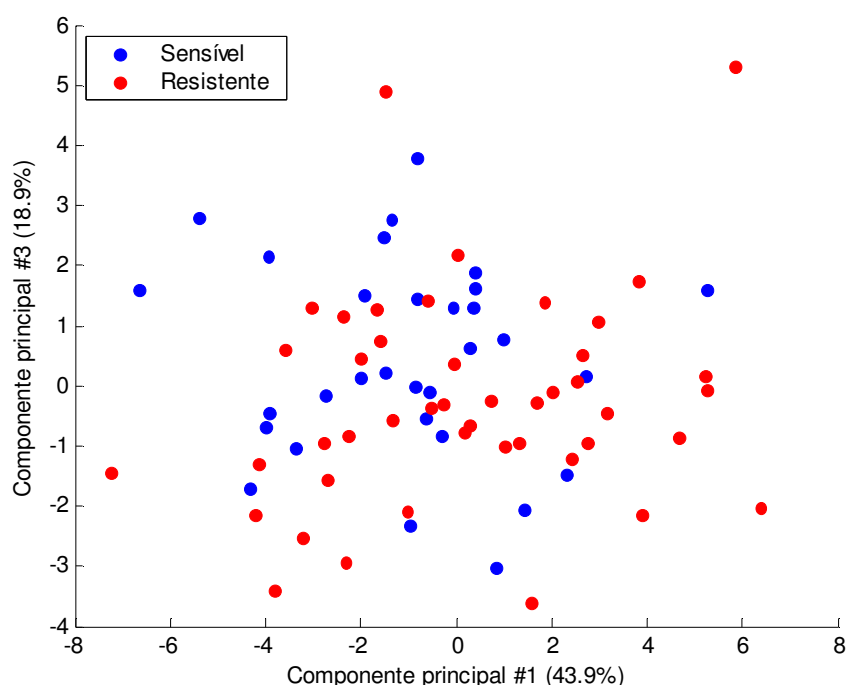


Figura III.59 - Análise de componentes principais para estirpes de *S. pneumoniae* cujos ácidos gordos foram extraídos e submetidos a uma análise por GC. Para esta análise os resultados foram submetidos a um pré-processamento de remoção da média.

Tal como foi referido para a secção da serotipagem, observando a Tabela VIII.1 (anexo III) verifica-se que existem colunas onde apenas existem ácidos gordos que estão relacionados com a resposta total da estirpe e não com a susceptibilidade à penicilina. Assim, tal como para a serotipagem, estas colunas foram eliminadas dos estudos. Foram também retirados os espectros

considerados aberrantes pela estatística de Hotelling e pela estatística dos resíduos. Fazendo uma representação gráfica do modelo de PCA com os dados resultantes, o que se obtém é o representado na Figura III.60.

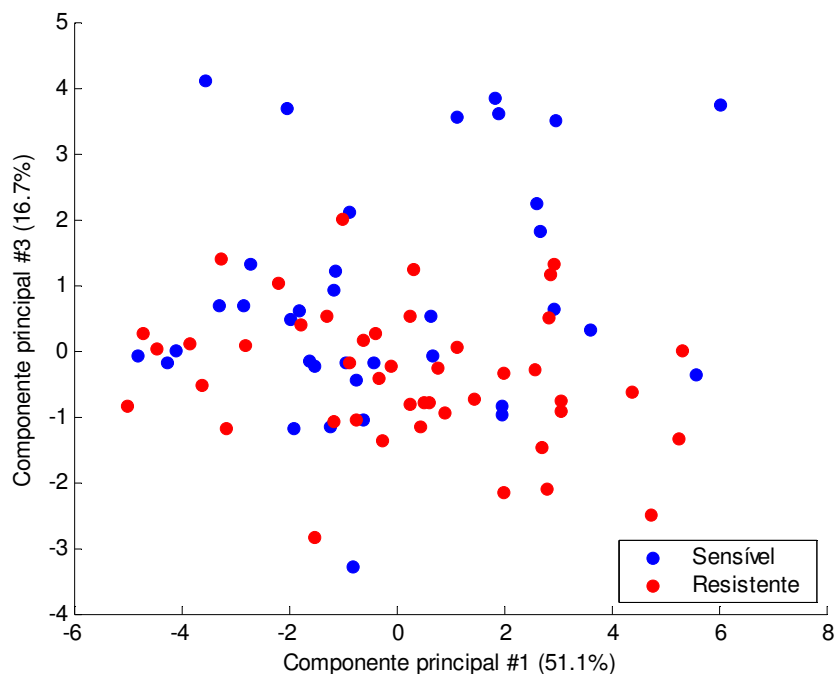


Figura III.60 - Análise de componentes principais para estirpes de *S. pneumoniae* cujos ácidos gordos foram extraídos e submetidos a uma análise por GC. Para esta análise os resultados foram submetidos a um pré-processamento de remoção da média e foram eliminados os resultados referentes aos resultados aberrantes. Foram também eliminadas da matriz as colunas referentes aos ácidos gordos em pequenas proporções.

Em seguida estudou-se o modelo SIMCA para o conjunto de dados da análise dos ácidos gordos de *S. pneumoniae* por GC aos quais foram retirados os dados aberrantes. Da matriz de resultados a analisar também foram retirados os ácidos gordos que apenas estão presentes nas estirpes cuja solução de análise estava mais concentrada, tal como descrito anteriormente. A representação gráfica final encontra-se na Figura III.61.

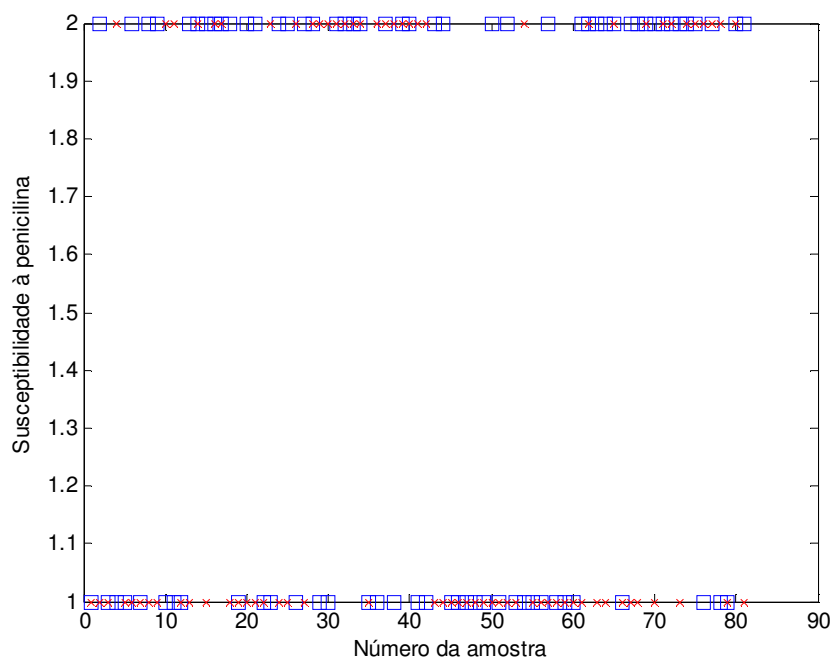


Figura III.61 - Resultado da validação cruzada para 3 blocos contíguos para o modelo SIMCA para estirpes de *S. pneumoniae* cujos ácidos gordos foram extraídos e submetidos a uma análise por GC. Para a construção do modelo foram retirados os resultados considerados aberrantes e o conjunto de resultados a analisar foi submetido à remoção da média. O número de componentes principais adequado é 7. Neste gráfico encontra-se representado no eixo das ordenadas a resistência à penicilina com o número 2 e a sensibilidade com o número 1. Representado com o símbolo (x) está a susceptibilidade prevista pelo modelo e com o símbolo (□) a susceptibilidade obtida pelo método CLSI. Há uma classificação incorrecta para 48,2% das estirpes.

O melhor modelo SIMCA para este conjunto de dados obtém-se para 7 componentes principais, mas a percentagem de resultados mal classificados é de 48,2%, o que não é aceitável. Dividindo os dados de modo a ter 70% para a construção do modelo e 30% para validar, a representação gráfica que se obtém dos dados de validação encontra-se na Figura III.62.

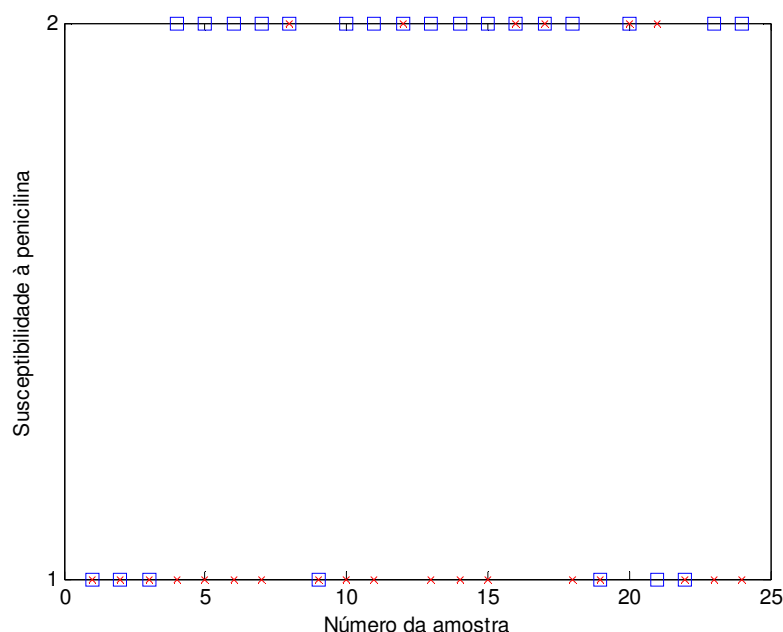


Figura III.62 - Resultado da validação do modelo SIMCA para 30% dos resultados da análise de GC para estirpes de *S. pneumoniae*. Para a construção do modelo foram usados os resultados que foram submetidos a um pré-processamento de remoção da média e foram usados 3 componentes principais. Para a construção do modelo foram retirados os resultados considerados aberrantes. Neste gráfico encontra-se representado no eixo das ordenadas a resistência à penicilina com o número 2 e a sensibilidade com o número 1. Representado com o símbolo (x) está a susceptibilidade prevista pelo modelo e com o símbolo (□) a susceptibilidade obtida pelo método CLSI. O modelo conseguiu prever correctamente 33,3% do resultado da susceptibilidade à penicilina.

O modelo SIMCA, para 30% dos resultados de GC usados na validação, conseguiu apenas prever correctamente a resistência/susceptibilidade à penicilina para 33,3%, o que não é aceitável para avaliar a susceptibilidade à penicilina de uma amostra desconhecida.

Testando o modelo do PLS discriminante para os dados descritos para a elaboração do modelo SIMCA, o melhor resultado que se obtém é para a representação da variável latente número 4 *versus* a variável latente número 1, que é o que se encontra representado na Figura III.63.

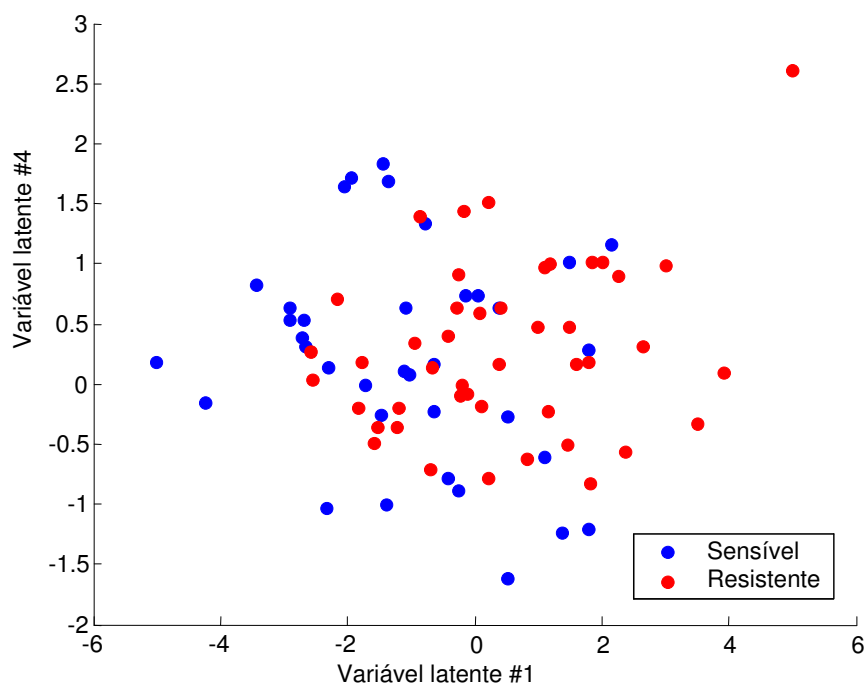


Figura III.63 - Resultados obtidos para o modelo de PLS discriminante usando para a validação cruzada 9 blocos contíguos e 8 variáveis latentes. Para a construção do modelo foram usados os resultados da extração dos ácidos gordos do *S. pneumoniae* que foram submetidos a um pré-processamento de remoção da média e que resultam da eliminação dos resultados de GC aberrantes e dos ácidos gordos sem resposta significativa. A equivalência entre as cores e a susceptibilidade encontra-se na figura.

Como se vê na figura acima representada, não existe uma boa separação entre as estirpes resistentes e sensíveis à penicilina. No entanto, construiu-se um modelo com 70% dos dados de GC que foi validado com os restantes 30%. O resultado da validação indica que o modelo de PLS discriminante consegue prever a susceptibilidade à penicilina para 62,5% dos resultados de GC.

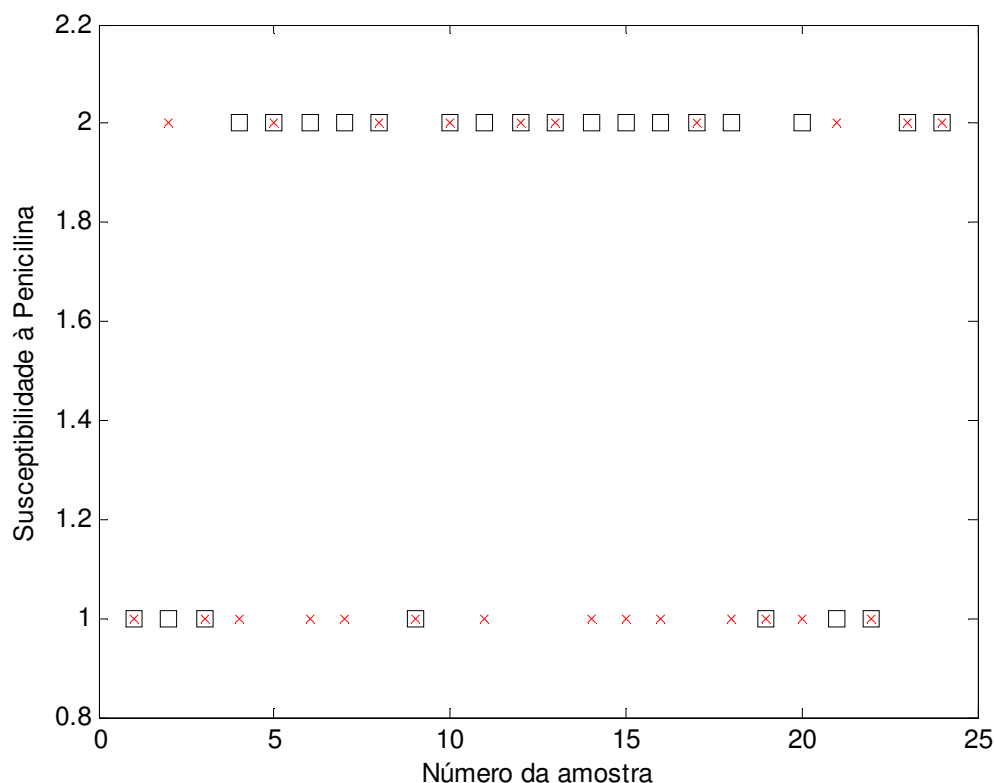


Figura III.64 - Resultados obtidos para a validação do modelo de PLS discriminante para 30% do conjunto de resultados da análise de ácidos gordos por GC de *S. pneumoniae*, dos quais foram eliminados os espectros aberrantes. Para a construção do modelo foram usadas 8 variáveis latentes. Os resultados foram submetidos à remoção da média e usaram-se apenas as colunas com quantidades de ácidos gordos significativas. Representada com o símbolo (x) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo método de referência. No eixo das ordenadas está a resistência à penicilina com o número 2 e a sensibilidade com o número 1. A percentagem de estirpes correctamente identificadas foi de 54,2%.

O modelo PLS discriminante conseguiu atribuir a 54,2% dos resultados de GC a resistência ou sensibilidade concordante com o método do CLSI.

Outro modelo supervisionado que pode ser testado é o PLS hierárquico. De modo a poder comparar os modelos foram usados os mesmos dados que para os modelos SIMCA e PLS discriminante. O modelo construído resulta de uma validação cruzada de 5 blocos contíguos e está descrito na Tabela III.11.

Tabela III.11- Resultados obtidos para o modelo de PLS hierárquico para os resultados da extracção dos ácidos gordos analisados por GC de *S. pneumoniae* resultantes da eliminação dos resultados aberrantes. Os resultados foram submetidos à remoção da média e foram eliminados os ácidos gordos sem expressão significativa. Nesta tabela estão duas linhas onde uma corresponde às estirpes resistentes e a outra às estirpes sensíveis cada uma com o número de variáveis latentes usadas para a identificação da susceptibilidade, a sensibilidade e a especificidade do modelo.

Teste de Susceptibilidade	LV	Sensibilidade (α) (%)	Especificidade (β) (%)
Sensível	6	61,11	73,33
Resistente	0	100,00	100,00

Tal como foi feito para os outros modelos, se se retirarem 70% dos resultados para a construção do modelo e se se validar este modelo com os 30% dos resultados restantes, o que se obtém é o representado na Figura III.65.

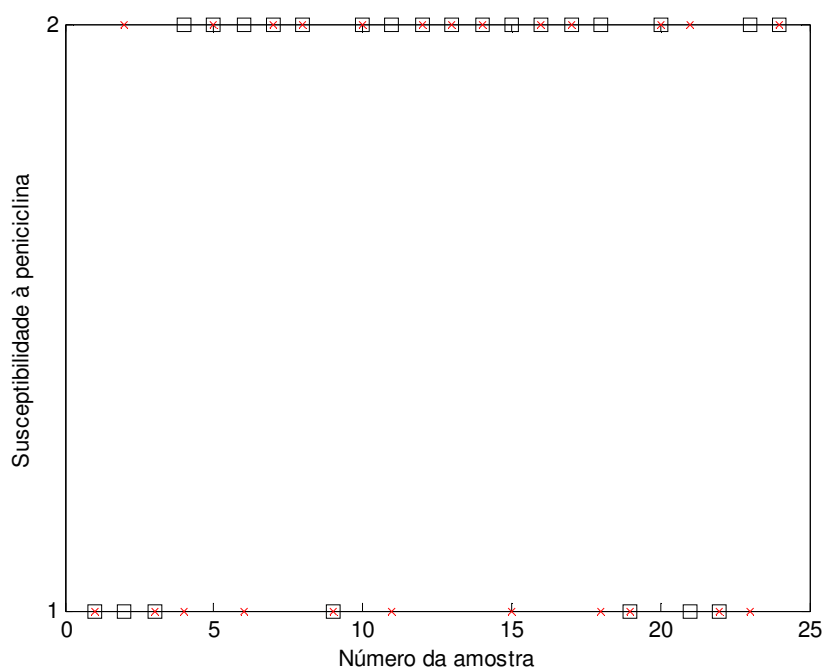


Figura III.65 - Validação do modelo PLS hierárquico com 30% do conjunto de resultados da análise de ácidos gordos por GC de *S. pneumoniae* dos quais foram eliminados os resultados aberrantes. Os resultados de GC foram submetidos a uma remoção da média e foram eliminados os ácidos gordos cuja expressão não é significativa. Neste gráfico encontram-se representados no eixo das ordenadas a sensibilidade à penicilina com o número 1 e a resistência com o número 2. Representada com o símbolo (x) está a susceptibilidade prevista pelo modelo e com o símbolo (□) a susceptibilidade definida pelo método descrito no CLSI. A taxa de previsões correctas é de 66,7%.

O modelo de PLS hierárquico apresentou uma taxa de resultados concordantes com o método do CLSI de 66,7%.

O último modelo supervisionado testado é o das redes neuronais. Se se usarem 4 nodos e 4 componentes principais, o resultado do modelo obtido e da validação é o representado na Figura III.66.

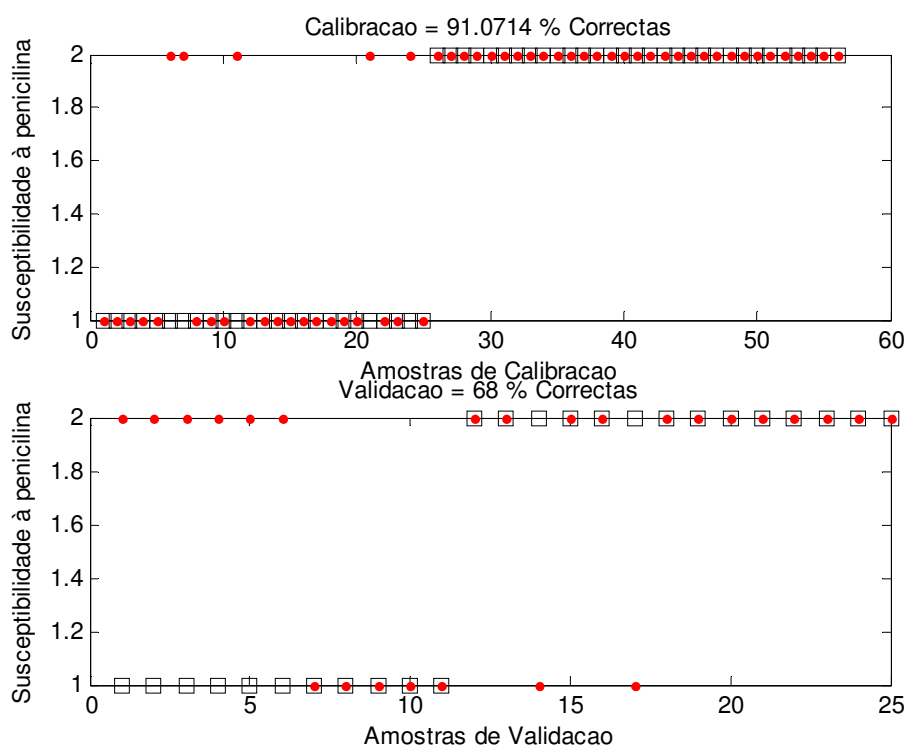


Figura III.66 - Representação gráfica dos resultados do modelo da rede neuronal aplicada aos resultados da análise de ácidos gordos por GC de *S. pneumoniae*. Foi usado como pré-processamento a remoção de média. Foram usados todos os ácidos gordos, 4 nodos e 4 componentes principais. Representada com o símbolo (●) está o serótipo previsto pelo modelo e com o símbolo (□) o serótipo definido pelo “método de xadrez”. No eixo das ordenadas estão os serótipos cuja equivalência é a seguinte: 1 corresponde ao serótipo 9V, o 2 ao serótipo 14, o 3 ao serótipo 19A e o 4 ao serótipo 23F.

Com as redes neuronais apenas se consegue atribuir a susceptibilidade à penicilina, concordante com os resultados obtidos pelo método do CLSI, para 68% dos ensaios.

Tal como para a serotipagem, não foi possível estabelecer uma relação entre a susceptibilidade à penicilina e os ácidos gordos de cadeia entre 9 e 20 átomos de carbono, pelas abordagens seguidas.

7. ATR

Na análise por ATR foi colocada uma colónia directamente no compartimento de amostras. Nesta técnica, o espectro de infra-vermelho resultante deverá ter uma contribuição da parede celular maior do que a do citoplasma, uma vez que o ATR analisa sobretudo a superfície das células^(70, 168). Assim, os resultados para a avaliação da serotipagem deverão ser melhores pois as diferenças nos serótipos acontecem ao nível da parede celular. Para o caso da susceptibilidade à penicilina, a resistência acontece pois as proteínas (PBP) responsáveis pela formação da parede celular, aquando da divisão celular, deixam de ter afinidade para a penicilina continuando o seu trabalho na presença ou ausência de antibiótico. Assim, a diferença entre estirpes resistentes e sensíveis deve-se à alteração de uma proteína existente na parede celular. Como as células têm muitas proteínas, os espectros de FT-IR obtidos são devidos às contribuições de todos os componentes celulares e onde todas as proteínas são agrupadas logo, terá uma grande vantagem se se conseguirem isolar as enzimas a estudar. Com o ATR não se isolam as enzimas mas, ao estudar apenas a parede celular, há um aumento da proporção das proteínas deste componente celular.

No ATR do equipamento de FT-IR, coloca-se a amostra no suporte de amostras, fecha-se o compartimento e aperta-se até onde se conseguir, não havendo nenhum equipamento acessório que controle a torção. Apenas foi testada a estirpe 2003V1490\$su19A uma vez que se queria avaliar a diferença

nos resultados entre as duas técnicas de amostragem: ATR e roda de selenato de zinco. O resultado obtido encontra-se na Figura III.67.

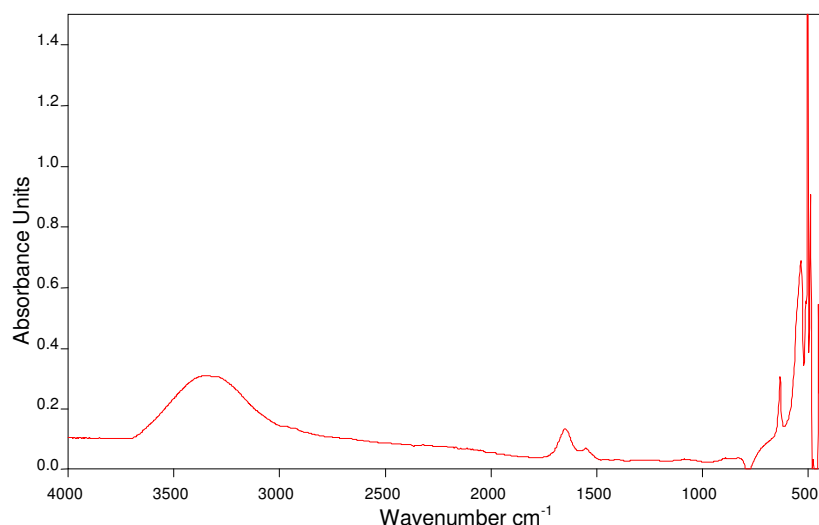


Figura III.67 - Espectro de ATR para a estirpe 2003V1490\$su19A de *S. pneumoniae*.

Como se pode observar da Figura III.67, o espectro resulta de uma quantidade insuficiente de amostra. No entanto, foi colocada não uma só colónia mas todo o conteúdo de uma placa de gelose de sangue. As colónias de pneumococos são muito pequenas, e para uma correcta avaliação da estirpe dever-se-íam preparar várias placas de gelose de sangue. Por outro lado, como é um estudo para elaboração de um modelo, dever-se-ia poder controlar a torção do ATR para garantir a reprodutibilidade da técnica. Por estas razões não foram realizados mais ensaios usando o acessório de ATR.

IV. CONCLUSÃO

Nesta dissertação descreve-se a aplicação da técnica de FT-IR a estirpes de *Streptococcus pneumoniae* com o objectivo de desenvolver um método de serotipagem para 4 serótipos (9V, 14, 19A e 23F) e um método para a avaliação da susceptibilidade à penicilina.

A técnica de FT-IR, como método para identificação de estirpes microbiológicas, é muito fácil de executar, é pouco dispendiosa e os resultados são bastante rápidos. Os consumíveis necessários para a análise de FT-IR fazem parte dos consumíveis habituais dos laboratórios e são bastante económicos. O único reagente necessário é água esterilizada. Assim, esta é uma técnica onde o valor gasto no ensaio de cada amostra é muito baixo, sobretudo quando comparado com os consumíveis e reagentes usados nos métodos de referência para a serotipagem e para os testes de susceptibilidade à penicilina. Uma vez que as estirpes de *S. pneumoniae* formam colónias pequenas e a quantidade de massa celular necessária para a análise em triplicado é grande, devem-se preparar duas placas de crescimento por cada estirpe. Como as placas usadas já preparadas foram adquiridas a um fornecedor, este será o consumível com maior importância ao nível económico.

Inicialmente a taxa de estirpes que passou o teste de qualidade definido pelo fabricante do equipamento de FT-IR (Bruker) foi baixa. No entanto, a técnica foi afinada com a prática e a taxa de sucesso foi aumentando, deixando de ser uma condicionante para a obtenção de resultados aceitáveis.

Para as estirpes analisadas foi possível construir modelos SIMCA, PLS discriminante, PLS hierárquico e redes neuronais com bons resultados

para a identificação dos serótipos 9V, 14,19A e 23F. Os modelos desenvolvidos pelas três técnicas apresentam resultados de identificação dos serótipos concordantes com o “método de xadrez”. Os resultados da correcta identificação das estirpes usadas na validação dos modelos são semelhantes, tal como descrito na Tabela IV.1.

Tabela IV.1 – Resumo das percentagens de correcta classificação dos espectros de FT-IR de acordo com o serótipo para os 4 modelos testados nesta dissertação.

Modelo	Construção do modelo (%)	Validação do Modelo (%)
SIMCA	93,5	91,8
PLS discriminante	92,6	97,1
PLS hierárquico	92,2	72,8
Redes neuronais	97,6	92,1

Assim, para a serotipagem dos pneumococos deve-se usar como modelo de quimiometria o PLS discriminante tendo em consideração que poderá haver um erro de classificação para 3% das estirpes. O único modelo que não deverá ser usado é o do PLS hierárquico.

De acordo com o descrito na validação do “método de xadrez”, quando as estirpes são reanalisadas por técnicos diferentes os resultados são todos concordantes, no entanto existem estirpes que não conseguem ser serotipadas, ou por não terem cápsula ou pelo serótipo não estar incluído no teste ⁽¹⁵⁵⁾. Num estudo realizado em 11 laboratórios de referência, em que se recorreu a métodos serológicos para serotipar *S. pneumoniae*, concluiu-se que houve um erro de 5% (39 serotipagens) num total de 735 serotipagens realizadas, onde 18% dos erros ocorreram ao nível da espécie, 28% ao nível do serogrupo ou no serótipo quando não pertencente a nenhum serogrupo, 44% acertou no serogrupo mas não no serótipo e para 10% foi registado que o

serótipo não era serotipável. Relativamente à percentagem tendo em conta apenas o número de estirpes, identificou-se um erro na serotipagem para 22 das 70 estirpes (305)⁽⁸³⁾. Assim, verifica-se que os métodos serológicos também estão associados a erros, e superiores aos descritos nesta tese. Para além dos erros, o método serológico é muito dispendioso pois implica a existência de muitos soros, implica muito trabalho, e técnicos formados para o efeito, ao contrário do método apresentado nesta dissertação.

Para a técnica de “dot-blot” a especificidade conseguida foi de 100% e a sensibilidade de 96%, valor próximo do obtido para o modelo PLS discriminante neste trabalho. A técnica é laboriosa e implica a existência de soros, tal como a técnica clássica, o que não é necessário para a análise de espectros de FT-IR⁽⁴²⁾. Os resultados da utilização do “Denka kit indicam uma especificidade e sensibilidade de 100% para a identificação de serogrupos mas com resultados piores para a identificação de serótipos ⁽¹⁰³⁾. O método descrito nesta dissertação é mais barato que o do referido kit, pois sendo um teste serológico acaba por necessitar de soros específicos que são sempre dispendiosos.

Outra técnica usada para serotipagem é o PCR. Numa publicação de 2003⁽²⁶⁾, foi descrita esta técnica em complemento com soros para identificar determinados serótipos. Assim, foram definidos iniciadores de DNA que definem grupos de serótipos e uma vez definido o grupo por PCR podem-se recorrer a novas reacções de PCR mas agora específicas para um dado serótipo ou recorrer a soros específicos. As técnicas de PCR são dispendiosas e implicam um trabalho muito controlado para evitar contaminações de DNA.

Para além disso a técnica de PCR descrita recorre a soros para identificar alguns serótipos, o que torna o método ainda mais dispendioso.

A técnica de FT-IR descrita neste trabalho é aquela que apresenta custos mais baixos por amostra analisada e que está associada a uma percentagem de erro muito baixa. No entanto, a limitação deste estudo é o número de serótipos estudados: apenas foram analisadas estirpes pertencentes a 4 serótipos identificados e um quinto grupo de estirpes cujo método de referência não identificou o serótipo. No que se refere à epidemiologia das crianças portuguesas com doença invasiva, a análise de estirpes dos serótipos 6B e 19F teria sido importante; no que refere a indivíduos com mais de 60 anos, o serótipo dominante encontra-se neste estudo mas existem outros serótipos com expressão significativa que não foram estudados: 1,3,4 e 8.

Para a avaliação da susceptibilidade à penicilina das estirpes de *S. pneumoniae*, não foi possível desenvolver métodos SIMCA, PLS discriminante, PLS hierárquico e de redes neuronais que conseguissem distinguir as estirpes resistentes das estirpes sensíveis. Como a resistência à penicilina está relacionada com a alteração da enzima PBP, as zonas do espectro que poderiam estar alteradas nos espectros das estirpes resistentes são as das proteínas e, nos espectros, estas estão divididas em várias zonas: amida I, amida II e amida III. Como estas zonas não são específicas de nenhuma proteína mas sim do efeito cumulativo de todas as proteínas na célula, é difícil observar a diferença do efeito de uma só enzima. Existem artigos que descrevem que foi possível distinguir estirpes resistentes de estirpes sensíveis a partir de espectros de FT-IR^(24, 51, 150). No entanto, nestes

casos, a resistência à penicilina não é devida à alteração de uma enzima mas sim, à expressão de proteínas que não são necessárias no caso de ausência de antibiótico. Nestes casos o efeito é visível e há uma sobre-expressão quando as estirpes crescem num meio com antibiótico⁽¹⁷⁷⁾. O ATR foi testado neste trabalho apenas para uma estirpe pois verificou-se que a quantidade de colónias de *S. pneumoniae* necessária para a obtenção de um espectro é muito grande, sendo para isso necessária a preparação de muitas placas de isolamento. Para além disso, o ATR disponível no laboratório do INFARMED, I. P., não tem controlo de torção, o que prejudica a repetibilidade do método que é importante aquando da construção de um modelo que se pretende que seja de uso fácil e aplicável por qualquer técnico, a qualquer altura. Por estes últimos factos a técnica de ATR deixou de ser testada.

Neste trabalho, encontra-se também descrita uma técnica de análise de ácidos gordos, por cromatografia gasosa, de estirpes de *S. pneumoniae*. O objectivo deste estudo foi avaliar a existência de uma relação entre a composição em ácidos gordos das estirpes e o seu serótipo ou a sua susceptibilidade à penicilina. Os resultados foram submetidos a um tratamento de análise multivariada e, nas condições deste estudo, não foi possível identificar uma relação entre ácidos gordos de cadeia entre C9 e C20 com os serótipos ou com a sensibilidade/resistência à penicilina.

V. PERSPECTIVAS FUTURAS

Relativamente ao serótipo, neste estudo apenas se testaram 4 serótipos de *S. pneumoniae* e um conjunto de estirpes pertencente a uma classe que não têm quantidade de polissacáridos capsulares suficientes ou cujo serótipo não se encontra representado no “método de xadrez”. Assim, de modo a completar este estudo, deveriam ser incluídas mais estirpes de outros serótipos, nomeadamente os mais frequentes na população com idade superior a 60 anos, que diferem dos serótipos analisados, e os serótipos 6B e 19F significativos em crianças com doença invasiva. A técnica usada pressupõe uma elevada quantidade de colónias que poderá ser reduzida se se utilizar um microscópio de FT-IR, neste caso seria apenas necessária uma microcolónia e a serotipagem poderia ser feita numa fase muito precoce do isolamento do microrganismo, diminuindo o número de dias de espera de um resultado. Também se poderiam usar outras técnicas tais como a espectroscopia de Raman, descrita também em muitos artigos com o intuito de identificar microrganismos. Seria também importante, para a consolidação dos modelos gerados, a criação de uma base de dados com as estirpes analisadas com o objectivo de identificar o serótipo de estirpes em rotina.

Relativamente ao estudo da susceptibilidade à penicilina do *S. pneumoniae*, não se obtiveram resultados conclusivos. A sensibilidade à penicilina deve-se ao facto das enzimas que participam na construção da parede celular (PBPs) se ligarem ao antibiótico impedindo a formação de novas células. Nas estirpes resistentes há uma alteração das PBPs que origina um decréscimo de afinidade ao antibiótico, ficando as enzimas disponíveis para a construção da parede celular. Logo, deve-se estudar esta zona do espectro

mas, como as proteínas existentes numa célula são muitas, as alterações nas PBPs podem não ser visíveis quando se estudam as contribuições de todos os componentes celulares. Assim, poder-se-iam submeter as estirpes, depois de crescidas em gelose de sangue, a um pré-tratamento que extraísse apenas a parede celular e fazer espectros apenas deste extracto tal como já foi referido numa publicação de 2005 ⁽¹⁷³⁾, ou utilizar um ATR com controlo de torção.

Por todas as vantagens descritas da técnica de FT-IR (rapidez na obtenção do resultado, custo por análise muito baixo, necessidade de poucos consumíveis, fácil de executar e sem necessidade de um operador experiente) e pelos resultados obtidos nesta dissertação, considera-se que esta técnica poderá ser implementada em laboratórios de rotina, como alternativa aos métodos clássicos de serotipagem.

VI. BIBLIOGRAFIA

1. 2002. Sherlock Microbial Identification System, Versão 4.5. MIDI, Newark, USA.
2. **Al-Quadiri, H. M., M. A. Al-Holy, M. Lin, N. I. Alami, A. G. Cavinato, e B. A. Rasco.** 2006. Rapid detection and identification of *Pseudomonas aeruginosa* and *Escherichia coli* as pure and mixed cultures in bottled drinking water using Fourier transform infrared spectroscopy and multivariate analysis. Journal of Agricultural and Food Chemistry **54**:5749-5754.
3. **Al-Tawfiq, J. A.** 2006. Antibiotic resistance of pediatric isolates of *Streptococcus pneumoniae* in a Saudi Arabian hospital from 1999 to 2004. Medical Science Monitor **12**:CR471-475.
4. **Amiali, N. A., M. R. Mulvey, J. Sedman, A. E. Simor, e A. A. Ismail.** 2007. Epidemiological typing of methicilin-resistant *Staphylococcus aureus* strains by Fourier transform infrared spectroscopy. Journal of Microbiological Methods **69**:146-153.
5. **Amiali, N. M., M. R. Mulvey, J. Sedman, M. Louie, A. E. Simor, e A. A. Ismail.** 2006. Rapid Identification of coagulase-negative staphylococci by Fourier transform infrared spectroscopy. Journal of Microbiological Methods **68**:236:242.
6. **Amiel, C., L. Mariey, C. Denis, P. Pichon, e J. Travert.** 2001. FTIR spectroscopy and taxonomic purpose: contribution to the classification of lactic acid bacteria. Lait **81**:249-255.

7. **Amiel, C., L. Mariey, M. C. Curk-Daubie, P. Pichon, e J. Travert.** 2000. Potentiality of Fourier transform infrared spectroscopy (FTIR) for discrimination and identification of dairy lactic acid bacteria. *Lait* **80**:445-459.
8. **Assunção, P., N. T. Antunes, R. S. Rosales, C. de la Fe, C. Poveda, J. B., Poveda, e H. M. Davey.** 2006. Flow cytometric method for the assessment of the minimal inhibitory concentrations of antibacterial agents to *Mycoplasma agalactiae*. *Cytometry A* **69**:1071-1076.
9. **Backhaus, E., S. Berg, B. Trollfors, R. Andersson, E. Persson, B .E. Claesson, P. Larsson, E. Ek, L. Jonsson, G. Radberg, S. Johansson, T. Ripa, D. Karlsson , e K. Andersson.** 2007. Antimicrobial susceptibility of invasive pneumococcal isolates from a region in south-west Sweden 1998-2001. *Scandinavian Journal of Infectious Diseases* **39**:19-27.
10. **Baldauf, N. A., L. A. Rodriguez-Romo, A. E. Yousef, e L. E. Rodriguez-Saona.** 2006. Differentiation of selected *Salmonella enterica* serovars by Fourier transform mid-infrared spectroscopy. *Applied Spectroscopy* **60**:592-598.
11. **Baldauf, N. A., L. A. Rodriguez-Romo, A. Mannig, A. E. Yousef, e L. E. Rodriguez-Saona.** 2007. Effect of selective growth media on the differentiation of *Salmonella enterica* serovars by Fourier-transform mid-infrared spectroscopy. *Journal of Microbiological Methods* **68**:106-114.
12. **Basheer, I. A., e M. Hajmeer.** 2000. Artificial neural networks: fundamentals, cumputing, design, and application. *Journal of Microbiological Methods* **43**:3-31.

13. **Bastert, J., H. C. Korting, P. Traenkle, e A. F. Schmalreck.** 1999. Identification of dermatophytes by Fourier transform infrared spectroscopy (FT-IR). *Mycoses* **42**:525-528.
14. **Beattie, S. H., C. Holt, D. Hirst, e A. G. William.** 1998. Discrimination among *Bacillus cereus*, *B. mycoides* and *B. thuringiensis* and some other species of the genus *Bacillus* by Fourier transform infrared spectroscopy. *FEMS Microbiology Letters* **164**:201-206.
15. **Becker, K., N. A. Laham, W. Fegeler, R. A. Proctor, G. Peters, e C. Von Eiff.** 2006. Fourier-transform infrared spectroscopic analysis is a powerful tool for studying the dynamic changes in *Staphylococcus aureus* small-colony variants. *Journal of Clinical Microbiology* **44**:3274-3278.
16. **Benbachir, M., S. Benredjeb, C. S. Boye, M. Dosso, H. Belabbes, A. Kamoun, O. Kaire, e N. Elmdaghri.** 2001. Two-year surveillance of antibiotic resistance in *Streptococcus pneumoniae* in four African cities. *Antimicrobial Agents Chemotherapy* **45**:627-629.
17. **Bergman, M., S. Huikko, P. Huovinen, P. Paakkari, H. Seppala, e Finnish Study Group for Antimicrobial Resistance (FiRe Network).** 2006. Macrolide and azithromycin use are linked to increased macrolide resistance in *Streptococcus pneumoniae*. *Antimicrobial Agents Chemotherapy* **50**:3646-3650.
18. **Beutin, L., Q. Wang, D. Naumann, W. Han, G. Krause, L. Leomil, L. Wang, e L. Feng.** 2007. Relationship between O-antigen subtypes, bacterial surface structures and O-antigen gene clusters in *Escherichia*

- coli*O123 strains carrying genes for Shiga toxin and intimin. Journal of Medical Microbiology **56**:177-184.
19. **Blossom, D. B., G. Namayanja-Kaye, J. Nankya-Mutyoba, J. B. Mukasa, H. Bakka, S. Rwambuya, A. Windau, S. Bajaksouzian, C. J. Walker, M. L. Joloba, C. Kityo, P. Mugenyi, C. C. Whalen, M. R. Jacobs, e R. A. Salata.** 2006. Oropharyngeal colonization by *Streptococcus pneumoniae* among HIV-infected adults in Uganda: assessing prevalence and antimicrobial susceptibility. International Journal of Infectious Diseases **10**:458-464.
20. **Bogaert, D., G. A. Syrogiannopoulos, I. N. Grivea, R. de Groot, N. G. Beratis, e P. W. Hermans.** 2000. Molecular epidemiology of penicillin-nonsusceptible *Streptococcus pneumoniae* among children in Greece. Journal of Clinical Microbiology **38**:4361-4366.
21. **Borel, M., e B. M. Lynch.** 1993. A study of living bacteria by ATR/FTIR spectroscopy. Canadian Journal of Applied Spectroscopy **38**:18-21.
22. **Borg, M., E. Scicluna, M. De Kraker, N. Van de Sande-Bruinsma, E. Tiemersma, D. Gür, S. B. Redjeb, O. Rasslan, Z. Elnassar, M. Benbachir, D. P. Bagatzouni, K. Rahal, Z. Daoud, H. Grundmann, e J. Monen.** 2006. Antibiotic resistance in the southeastern Mediterranean - preliminary results from the ARMed project. Euro Surveillance **11**:164-167
disponível na página:
<http://www.eurosurveillance.org/em/v11n07/1107-226.asp>.
23. **Bosch, A., M. A. Golowczyc, A. G. Abraham, G. L. Garrote, G. L. De Antoni, e O. Yantorno.** 2006. Rapid discrimination of lactobacilli isolated

- from kefir grains by FT-IR spectroscopy. *International Journal of Food Microbiology* **111**:280-287.
24. **Bouhedja, W., G. D. Sockalingum, P. Pina, P. Allouch, C. Bloy, R. Lábia, J. M. Millot, e M. Manfait.** 1997. ATR-FTIR spectroscopic investigation of *E. coli* transconjugants b-lactams-resistance phenotype. *FEBS Letters* **412**:39-42.
25. **Brandileone, M. C., J. L. Di Fabio, V. S. Vieira, R. C. Zanella, S. T. Casagrande, A. C. Pignatari, e A. Tomasz** 1998. Geographic distribution of penicillin resistance of *Streptococcus pneumoniae* in Brazil: genetic relatedness. *Microbial Drug Resistance* **4**:209-217.
26. **Brito, D. A., M. Ramirez, e H. Lencastre.** 2003. Serotyping *Streptococcus pneumoniae* by multiplex PCR. *Journal of Clinical Microbiology* **41**:2378-2384.
27. **Bruker.** Working with the IFS 28/B and B-module. Sample preparation, measurement and evaluation of the FT-IR spectra of microorganisms, Preliminary ed. Bruker.
28. **Burgula, Y., D. Khali, S. Kim, S. S. Krishnan, M. A. Cousin, J. F. Gore, H. L. Reuhs, e L. J. Mauer.** 2006. Detection of *Escherichia coli* O157:H7 and *Salmonella thyphimurium* using filtration followed by Fourier-transform infrared spectroscopy. *Journal of Food Protection* **69**:2777-2784.
29. **Camargos, P., G. B. Fischer, H. Mocelin, C. Dias, e R. Ruvinsky.** 2006. Penicillin resistance and serotyping of *Streptococcus pneumoniae* in Latin America. *Paediatric Respiratory Reviews* **7**:209-214.

30. **Castanheira, M., R. N. Jones, S. Silbert, M. C. Brandileone, e H. S. Sader.** 2003. Typing and molecular characterization of *Streptococcus pneumoniae* with reduced susceptibility to cefotaxime isolated in Latin America. *Microbial Drug Resistance* **9**:345-351.
31. **Chalmers, J. M., e P.R. Griffiths.** 2002. Handbook of Vibrational Spectroscopy. John Wiley & Sons, Chichester.
32. **Chi, F., O. Nolte, C. Bergmann, M. Ip, e R. Hakenbeck.** 2007. Crossing the barrier: evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. *International Journal of Medical Microbiology* **297**:503-512.
33. **Clinical and Laboratory Standards Institute.** 2007. Performance standards for antimicrobial susceptibility testing; seventeenth informational supplement, CLSI document M100-S17, West Valey Road, Pensylvania, USA.
34. **Curk, M. C., F. Peladan, e J. C. Hubert.** 1994. Fourier-transform infrared (FTIR) spectroscopy for identifying lactobacillus species. *FEMS Microbiology Letters* **123**:241-248
35. **de Neeling, A. J., B. P. Overbeek, A. M. Horrevorts, E. E. Ligtoet, e W. G. Goettsch.** 2001. Antibiotic use and resistance of *Streptococcus pneumoniae* in The Netherlands during the period 1994-1999. *Journal of Antimicrobial Chemotherapy* **48**:441-444.
36. **del Campo, R., F. Cafini, M. I. Morosini, A. Fenoll, J. Linares, L. Alou, D. Sevillano, R. Canton, J. Prieto, F. Baquero, e Spanish Pneumococcal Network (G3/103).** 2006. Combinations of PBPs and MurM protein variants in early and contemporary high-level penicillin-

- resistant *Streptococcus pneumoniae* isolates in Spain. Journal of Antimicrobial Chemotherapy **57**:983-986.
37. **Dias, R., D. Louro, The antimicrobial Resistance Surveillance Program in Portugal, e M. Caniça.** 2006. Antimicrobial susceptibility of invasive *Streptococcus pneumoniae* isolates in Portugal over an 11-year period. Antimicrobial Agents Chemotherapy **50**:2098-2105.
38. **Dziuba, B., A. Babuchowski, D. Nałęcz, e M. Niklewicz.** 2007. Identification of lactic acid bacteria using FTIR spectroscopy and cluster analysis. International Dairy Journal **17**:183-189.
39. **EARSS.** 2006. EARSS Annual Report 2005.1-147, disponível na página <http://www.rivm.nl/EARSS>.
40. **Einax, J. W.** 2004. Chemometrics in analytical chemistry. Analytical and Bionalytical Chemistry **380**:368-369.
41. **Feikin, D. R., M. Davis, O. C. Nwanyanwu, P. N. Kazembe, L. M. Barat, A. Wasas, P. B. Bloland, C. Ziba, T. Capper, R. E. Huebner, B. Schwartz, K. P. Klugman, e S. F. Dowell.** 2003. Antibiotic resistance and serotype distribution of *Streptococcus pneumoniae* colonizing rural Malawian children. Pediatric Infectious Disease Journal **22**:564-567.
42. **Fenoll, A., I. Jado, D. Vicioso, e J. Casal.** 1997. Dot blot assay for the serotyping of pneumococci. Journal of Clinical Microbiology **35**:764-766.
43. **Garau, J.** 2005. Role of beta-lactam agents in the treatment of community-acquired pneumonia. European Journal of Clinical Microbiology & Infectious Diseases **24**:83-99.
44. **García-Suárez, M. d. M., R. Villaverde , A. F. Caldevilla , F. J. Mendéz , F. Vázquez.** 2006. Serotype distribution and antimicrobial resistance of

- invasive and non-invasive pneumococccal isolates in Asturias, Spain. Japanese Journal of Infectious Diseases **59**:299-305.
45. **Garip, S., F. Bozoglu, e F. Severcan.** 2007. Differentiation of mesophilic and thermophilic bacteria with Fourier transform infrared spectroscopy. Applied Spectroscopy **61**:186-192.
46. **Geladi, P., e B. Kowalski.** 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta **185**:1-17.
47. **Gherardi, G., L. Fallico, M. Del Grosso, F. Bonanni, F. D'Ambrosio, R. Manganelli, G. Palu, G. Dicuonzo, e A. Pantosti.** 2007 Antibiotic-resistant invasive pneumococcal clones in Italy. Journal of Clinical Microbiology **45**:306-312.
48. **Goodacre, R.** 2003. Explanatory analysis of spectroscopic data using machine learning of simple interpretable rules. Vibrational Spectroscopy **32**:33-45.
49. **Goodacre, R., E. M. Timmins, P. J. Rooney, J. J. Rowland, e D. B. Kell.** 1996. Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks. FEMS Microbiology Letters **140**:233-239.
50. **Goodacre, R., E. M. Timmins, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell, e P. J. Rooney.** 1998. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. Microbiology **144**:1157-1170.
51. **Goodacre, R., P. J. Rooney, e D. B. Kell.** 1998. Rapid Analysis of microbial systems using vibrational spectroscopy and supervised

- learning methods: application to the discrimination between methicillin-resistant and methicillin-susceptible *Staphylococcus aureus*, p. 220-227. In H. H. Mantsch, e M. Jackson (ed.), Infrared spectroscopy: new tool in medicine, vol. 3257. Proceedings of SPIE.
52. **Gosbell, I. B., L. A. Fernandes, e C. J. Fernandes.** 2006. In vitro antibacterial activity of beta-lactams and non-beta-lactams against *Streptococcus pneumoniae* isolates from Sydney, Australia. Pathology **38**:343-348.
53. **Greenberg, D., R. Dagan, M. Muallem, e N. Porat.** 2003. Antibiotic-resistant invasive pediatric *Streptococcus pneumoniae* clones in Israel. Journal of Clinical Microbiology **41**:5541-5545.
54. **Gupta, M. J., J. Irudayaraj, e C. Debroy.** 2004. Spectroscopic contamination of bacteria using artificial neural networks. Journal of Food Protection **67**:2550-2554.
55. **Gwanzura, L., C. Pasi, K. J. Nathoo, J. Hakim, I. Gangaidzo, J. Mielke, V. J. Robertson, R. S. Heyderman, e P. R. Mason.** 2003. Rapid emergence of resistance to penicillin and trimethoprim-sulphamethoxazole in invasive *Streptococcus pneumoniae* in Zimbabwe. International Journal of Antimicrobial Agents **21**:557-561.
56. **Haag, H., H. U. Gremlich, R. Bergmann, e J. J. Sanglier.** 1996. Characterization and Identification of Actinomycetes by FT-IR spectroscopy. Journal of Microbiological Methods **27**:157-163.
57. **Hakenbeck, R., H. Ellerbrok, T. Briese, S. Handwerger, e A. Tomasz.** 1986. Penicillin-binding proteins of penicillin-susceptible and -resistant pneumococci: immunological relatedness of altered proteins and

- changes in peptides carrying the beta-lactam binding site. *Antimicrobial Agents Chemotherapy* **30**:553-558.
58. **Hausdorff, W. P., G. Siber, e P. R. Paradiso.** 2001. Geographical differences in invasive pneumococcal disease rates and serotype frequency in young children. *Lancet* **357**:950-952.
59. **Helm, D., H. Labischinski, e D. Naumann.** 1991. Elaboration of a procedure for identification of bacteria using Fourier-transform infrared spectral libraries: a stepwise correlation approach. *Journal of Microbiological Methods* **14**:127-142.
60. **Helm, D., H. Labischinski, G. Schallehn, e D. Naumann.** 1991. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *Journal of General Microbiology* **137**:69-79.
61. **Henrichsen, J.** 1979. The pneumococcal typing system and pneumococcal surveillance. *Journal of Infectious Diseases* **1**:31-37.
62. **Hill, P. C., A. Akisanya, K. Sankareh, Y. B. Cheung, M. Saaka, G. Lahai, B. M. Greenwood, e R. A. Adegbola.** 2006. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers. *Clinical Infectious Diseases* **43**:673-679.
63. **Holt, C., D. Hirst, A. Sutherland, e F. MacDonald.** 1995. Discrimination of species in the genus *Listeria* by Fourier transform infrared spectroscopy and canonical variate analysis *Applied and Environmental Microbiology* **61**:377-378.
64. **Holt, J. G., N. R. Krieg, P. H. Sneath, J. T. Staley, e S. T. Williams.** 2000. *Bergeys's Manual of Determinative Bacteriology*, 9 ed. Lippincott Williams & Wilkins, Baltimore, USA.

65. **Holtze, M. S., P. Nielsen, F. Ekelund, L. D. Rasmussen, e K. Johnsen.** 2006. Mercury affects the distribution of culturable species of *Pseudomonas* in soil. *Applied soil Ecology* **31**:228-238.
66. **Horbach, I., D. Naumann, e F. J. Fehrenbach.** 1988. Simultaneous infections with different serogroups of *Legionella pneumophila* investigated by routine methods and Fourier transform infrared spectroscopy. *Journal of Clinical Microbiology* **26**:1106-1110.
67. **Ip, M., S. S. Chau, F. Chi, E. S. Cheuk, H. Ma, R. W. Lai, e P. K. Chan.** 2007. Longitudinally tracking fluoroquinolone resistance and its determinants in Penicillin - susceptible and non-susceptible *Streptococcus pneumoniae* in Hong Kong, 2000 to 2005. *Antimicrobial Agents Chemotherapy* **51**:2192-2194.
68. **Irmscher, H. M., R. Fischer, W. Beer, e G. Selmann.** 1999. Characterization of nosocomial *Serratia marcescens* isolates: comparison of Fourier-transform infrared spectroscopy with pulse-field gel electrophoresis of genomic DNA fragments and multilocus enzyme electrophoresis. *Zentralbl Bakteriologie* **289**:249-263.
69. **Isozumi, R., Y. Ito, T. Ishida, M. Osawa, T. Hirai, I. Ito, K. Maniwa, M. Hayashi, H. Kagioka, M. Hirabayashi, K. Onari, H. Tomioka, K. Tomii, I. Gohma, S. Imai, S. Takakura, Y. Iinuma, S. Ichiyama, M. Mishima, e Kansai Community Acquired Pneumococcal Pneumonia Study Groupe.** 2007. Genotypes and related factors reflecting macrolide resistance in pneumococcal pneumonia infections in Japan. *Journal of Clinical Microbiology* **45**:1440-1446.

70. **Jiang, W., A. Saxena, B. Song, B. B. ward, T. J. Beverige, e S. B. Myneni.** 2004. Elucidation of functional groups on Gram-positive and Gram-negative bacterial surfaces using infrared spectroscopy. *Langmuir* **20**:11433-11442.
71. **Johnsen, K., e P. Nielsen.** 1999. Diversity of *Pseudomonas* strains isolated with King's B and Gould's S1 agar determined by repetitive extragenic palindromic-polymerase chain reaction, 16S rDNA sequencing and Fourier transform infrared spectroscopy characterisation. *FEMS Microbiology Letters* **173**:155-162.
72. **Jorgensen, J. H., J. D. Turnidge, e J. A. Washington.** Antibacterial susceptibility tests: dilution and disk diffusion methods, p. 1526-1543. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover e R. H. Tenover (ed.), *Manual of Clinical Microbiology*, 7 ed. American Society for Microbiology, Washington.
73. **Joung, Y. H., H. R. Kim, M. K. Lee, e A. J. Park.** 2007. Fluconazole susceptibility testing of *Candida* species by flow cytometry. *Journal of Infection* **54**:504-508.
74. **Kansiz, M., P. Heraud, B. Wood, F. Burden, J. Beardall, e D. Mac Naughton.** 1999. Fourier transform infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains. *Phytochemistry* **52**:407-417.
75. **Katz, A., E. Leibovitz , V. N. Timchenko , D. Greenberg , N. Porat , N. Peled , R. Dagan , e I. B. Ossipov.** 2007. Antibiotic susceptibility, serotype distribution and vaccine coverage of nasopharyngeal and oropharyngeal *Streptococcus pneumoniae* in a day-care centre in St.

- Petersburg, Russia. *Scandinavian Journal of Infectious Diseases* **39**:293-298.
76. **Kim, S., B. L. Reuhs, e L. J. Mauer.** 2005. Use of Fourier transform infrared spectra of crude bacterial lipopolysaccharides and chemometrics for differentiation of *Salmonella enterica* serotypes. *Journal of Applied Microbiology* **99**:411-417.
77. **Kim, S., H. Kim, B. L. Reuhs, e L. J. Mauer.** 2006. Differentiation of outer membrane proteins from *Salmonella enterica* serotypes using Fourier transform infrared spectroscopy and chemometrics. *Letters in Applied Microbiology* **42**:229-234.
78. **Kirkwood, J., A. Ghetler, J. Sedman, D. Leclair, F. Pagotto, J. W. Austin, e A. A. Ismail.** 2006. Differentiation of group I and group II strains of *Clostridium botulinum* by focal plane array Fourier transform infrared spectroscopy. *Journal of Food Protection* **69**:2377-2383.
79. **Klugman, K. P.** 2002. The successful clone: the vector of dissemination of resistance in *Streptococcus pneumoniae*. *Journal of Antimicrobial Chemotherapy* **50**:1-5.
80. **Klugman, K. P., T. J. Coffey, A. Smith, A. Wasas, M. Meyers, e B. G. Sprat.** 1994. Cluster of an erythromycin-resistant variant of the Spanish multiply resistant 23F clone of *Streptococcus pneumoniae* in South Africa. *European Journal of Clinical Microbiology & Infectious Diseases* **13**:171-174.
81. **Kohanteb, J., e E. Sadeghi.** 2007. Penicillin-resistant *Streptococcus pneumoniae* in Iran. *Medical Principles and Practice* **16**:29-33.

82. **Konradsen, H. B., e M. S. Kalsoff.** 2002. Invasive pneumococcal infections in Denmark from 1995 to 1999: epidemiology, serotypes, and resistance. *Clinical and Diagnostic Laboratory Immunology* **9**:358-365.
83. **Konradsen, H. B., e Pneumococcus Reference Laboratories in Europe.** 2005. Validation of serotyping of *Streptococcus pneumoniae* in Europe. *Vaccine* **23**:1368-1373.
84. **Kourti, T.** 2002. Process analysis and abnormal situation detection: from theory to practice. *IEEE control Systems* **22**:10-25.
85. **Krishna, C. M., G. Kegelaer, I. ADT, S. Rubin, V. B. Kartha, M. Manfait, e G. D. Sockalingum.** 2006. Combined Fourier transform infrared and Raman spectroscopic approach for identification of multidrug resistance phenotype in cancer cell lines. *Biopolymers* **82**:462-470.
86. **Kümmerle, M., S. Scherer, e H. Seiler.** 1998. Rapid and reliable identification of food-borne yeasts by Fourier-transform infrared spectroscopy. *Applied and Environmental Microbiology* **64**:2207-2214.
87. **Lamprell, H., G. Mazerolles, A. Kodjo, F. F. Chamba, Y. Noël, e E. Beuvier.** 2006. Discrimination of *Staphylococcus aureus* strains from different species of *Staphylococcus* using Fourier transform infrared (FTIR) spectroscopy. *International Journal of Food Microbiology* **108**:125-129.
88. **Lang, P. L., e S. C. Sang.** 1998. The *in situ* infrared microspectroscopy of bacterial colonies on agar plates. *Cellular Molecular Biology* **44**:231-238.
89. **Lavine, B. K.** 1998. Chemometrics. *Analytical Chemistry* **70**:209R-228R.

90. **Lefevre, J. C., M. A. Bertrand, e G. Faucon.** 1995. Molecular analysis by pulsed-field gel electrophoresis of penicillin-resistant *Streptococcus pneumoniae* from Toulouse, France. *European Journal of Clinical Microbiology & Infectious Diseases* **14**:491-497.
91. **Lefier, D., D. Hirst, C. Holt, e A. G. Williams.** 1997. Effect of sampling procedure and strain variation in *Listeria monocytogenes* on the discrimination of species in the genus *Listeria* by Fourier transform infrared spectroscopy and canonical variates analysis. *FEMS Microbiology Letters* **147**:45-50.
92. **Lefier, D., H. Lamprell, e G. Mazerolles.** 2000. Evolution of *Lactococcus* strains during ripening in Brie cheese using Fourier transform infrared spectroscopy. *Lait* **80**:247-254.
93. **Lemmer, K., D. Naumann, B. Raddatz, e K. Tintelnot.** 2004. Molecular typing of *Cryptococcus neoformans* by PCR fingerprinting, in comparison with serotyping and Fourier transform infrared-spectroscopy-based phenotyping. *Medical Mycology* **42**:135-147.
94. **Lipkus, A. H., K. K. Chittur, S. J. Vesper, J. B. Rbinson, e G. E. Pierce.** 1990. Evaluation of infrared spectroscopy as a bacterial identification method. *Journal of Indian Microbiology* **6**:71-75.
95. **Livermore, D. M., R. Reynolds, P. Stephens, G. Duckworth, D. Felmingham, A. P. Johnson, S. Murchan, O. Murphy, U. Gungabissoon, P. Waight, R. Pebody, J. Shackcloth, M. Warner, L. Williams, e R. C. George.** 2006. Trends in penicillin and macrolide resistance among pneumococci in the UK and the Republic of Ireland in

- relation to antibiotic sales to pharmacies and dispensing doctors. *International Journal of Antimicrobial Agents* **28**:273-279.
96. **Lopes, J.** 2001. Supervisão e diagnóstico de processos farmacêuticos com métodos inteligentes de análise de dados. Universidade Técnica de Lisboa, Lisboa.
97. **Louwerse, D. J., e A. K. Smilde.** 2000. Multivariate statistical process controlo of batch processes based on three-way models. *Chemical Engineering Science* **55**:1225-1235.
98. **Maquelin, K., L.-P. Choo-Smith, C. Kirschner, N. A. Ngo-Thi, D. Naumann, e G. J. Puppels.** 2002. Vibrational spectroscopic studies of microorganisms, p. 3308-3334. *In* J. M. e. P. R. G. Chalmers (ed.), *Handbook of Vibrational Spectroscopy*, 1^a ed, vol. 5. John Wiley & Sons, Chichester.
99. **Mariey, L., J. P. Signolle, C. Amiel, e J. Travert.** 2001. Discrimination, classification, identification of microorganisms using FTIR Spectroscopy and chemometrics. *Vibrational Spectroscopy* **26**:151-159.
100. **Mato, R., I. S. Sanches, C. Simas, S. Nunes, J. A. Carriço, N. G. Sousa, N. Frazão, J. Saldanha, A. Brito-Avô, J. S. Almeida, e H. De Lencastre.** 2005. Natural history of drug-resistant clones of *Streptococcus pneumoniae* colonizing healthy children in Portugal. *Microbial Drug Resistance* **11**:309-322.
101. **Miller, C. E.** 2005. Chemometrics in process analytical chemistry. *In* K. A. Bakeev (ed.), *Process Analytical Technology*. Blackwell Publishing, Oxford, UK, 451pp.

102. **Miller, J. N., e J. C. Miller.** 2005. Statistics and chemometrics for analytical chemistry, 5 ed. Pearson Educational Limited, Inghilterra.
103. **Mudany, M. A., K. Kikuchi, K. Totsuka, e T. Uchiyama.** 2003. Evaluation of a new serotyping kit for *Streptococcus pneumoniae*. Journal of Medical Microbiology **52**:975-980.
104. **Murray, P. R., K. S. Rosenthal, G. S. Kobayashi, M. A. Pfaller.** 2002. Medical Microbiology, 4^a ed ed. Mosby inc St Louis Missouri.
105. **Næs, T., T. Isaksson, T. Fearn, e T. Davies.** 2004. Multivariate calibration and classification. NIR publications, Chichester, West Sussex, Inghilterra.
106. **Nau, R., e H. Eiffert.** 2002. Modulation of release of proinflammatory bacterial compounds by antibacterials: potential impact on course of inflammation and outcome in sepsis and meningitis. Clinical Microbiology Reviews **15**:95-110.
107. **Naumann, D.** 2001. FT-Infrared and FT-Raman spectroscopy in biomedical research. Applied Spectroscopy Reviews **36**:239-298.
108. **Naumann, D., D. Helm, e H. Labischinski.** 1991. Microbiological characterizations by FT-IR spectroscopy. Nature **351**:81-82.
109. **Naumann, D., V. Fijala, H. Labischinski, e P. Giesbrecht.** 1988. The differentiation and identification of pathogenic bacteria using FT-IR and multivariate statistical analysis. Mikrochimica Acta **1**:373-377.
110. **Naumann, D., V. Fijala, H. Labischinski, e P. Giesbrecht.** 1988. The rapid differentiation and identification of pathogenic bacteria using Fourier transform infrared spectroscopic and multivariate statistical analysis. Journal of Molecular Structure **174**:165-170.

111. **Neugebauer, U., U. Schmid, K. Baumann, W. Ziebuhr, S. Kozitskaya, V. Deckert, M. Schmitt, e J. Popp.** 2007. Towards a detailed understanding of bacterial metabolism - spectroscopic characterization of *Staphylococcus epidermidis*. *ChemPhysChem* **8**:124-127.
112. **Neves, H. J. C., e A. M. C. Freitas.** 1996. Introdução à cromatografia gás-líquido de alta resolução. Dias de Sousa, Lda, Póvoa de Santa Iria, Portugal.
113. **Ngo Thi, N. A., C. Kirschner, e D. Naumann.** 2003. Characterization and Identification of microorganisms by FT-IR microspectrometry. *Journal of Molecular Structure* **661-662**:371-380.
114. **Ngo Thi, N. A., e D. Naumann.** 2006. Investigating the heterogeneity of cell growth in microbial colonies by FTIR microspectroscopy. *Analytical and Bionalytical Chemistry* **387**:1769-1777.
115. **Norris, K. P.** 1959. Infrared spectroscopy and its application to microbiology. *The Journal of Hygiene* **57**:326-345.
116. **Oberreuter, H., F. Mertens, H. Seiler, e S. Scherer.** 2000. Quantification of micro-organism in binary mixed populations by Fourier transform infrared (FT-IR) spectroscopy. *Letters in Applied Microbiology* **30**:85-89.
117. **Orrett, F. A., e E. Changoor.** 2007. Bacteremia in children at a regional hospital in Trinidad. *International Journal of Infectious Diseases* **11**:145-151.
118. **Otto, M.** 2007. Chemometrics - statistics and computer application in analytical chemistry, 2ª ed. Wiley-VCH, Weinheim.

119. **Palhares, R., F. Gudiol, J. Linares, J. Ariza, G. Rufi, L. Murgui, J. Dorca, e P. F. Viladrich.** 1987. Risk factors and response to antibiotic therapy in adults with bacteremic pneumonia caused by penicillin-resistant pneumococci. *New England Journal of Medicine* **317**:18-22.
120. **Park, I. H., D. G. Pritchard , R. Cartee , A. Brandao , M. C. Brandileone , e M. H. Nahm.** 2007. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology* **45**:1225-1233.
121. **Pato, M. V., C. B. Carvalho, A. Tomasz, e Multicenter Study Group.** 1995. Antibiotic susceptibility of *Streptococcus pneumoniae* isolates in Portugal. A multicenter study between 1989 and 1993. *Microbial Drug Resistance* **1**:59-69.
122. **Preisner, O., J. A. Lopes, R. Guiomar, J. Machado, e J. C. Menezes.** 2006. Fourier transform infrared (FT-IR) spectroscopy in bacteriology: towards a reference method for bacteria discrimination. *Analytical and Bioanalytical Chemistry*.
123. **Quinones-Falconi, F., J. J. Calva, Y. Lopez-Vidal, M. Galicia-Velazco, M. E. Jimenez-Martinez, e L. Larios-Mondragon.** 2004. Antimicrobial susceptibility patterns of *Streptococcus pneumoniae* in Mexico. *Diagnostic Microbiology Infectious Disease* **49**:53-58.
124. **Rebuffo-Scheer, C. A., C. Kirschner, M. Staemmler, D. Naumann.** 2007. Rapid species and strain differentiation of non-tuberculous mycobacteria by Fourier-transform infrared microspectroscopy. *Journal of Microbiological Methods* **68**:282-290.

125. **Rebuffo-Scheer, C. A., J. Schimtt, e S. Scherer.** 2007. Differentiation of *Listeria monocytogenes* serovars by using artificial neural network analysis of Fourier-transformed infrared spectra. *Applied and Environmental Microbiology* **73**:1036-1040.
126. **Reichmann, P., E. Varon , E. Gunther, R. R. Reinert , R. Luttiken , A. Marton , P. Geslin , J. Wagner, e R. Hakenbeck.** 1995. Penicillin-resistant *Streptococcus pneumoniae* in Germany: genetic relationship to clones from other European countries. *Journal of Medical Microbiology* **43**:377-385.
127. **Reinert, R. R., M. Van Der Linden, I. Seegmuller, A. Al-Lahham, A. Siedler, B. Weissmann, A. M. Toschke, e R. Von Kries.** 2007. Molecular epidemiology of penicillin-non-susceptible *Streptococcus pneumoniae* isolates from children with invasive pneumococcal disease in Germany. *Clinical Microbiology and Infection* **13**:363-368.
128. **Reis, E.** 2001. Estatística multivariada aplicada, 1 ed. Edições Sílabo, Lisboa.
129. **Richter, S. S., K. P. Heilmann, S. L. Coffman, H. K. Huynh, A. B. Brueggemann, M. A. Pfaller, e G. V. Doern.** 2002. The molecular epidemiology of penicillin-resistant *Streptococcus pneumoniae* in the United States, 1994-2000. *Clinical Infectious Diseases* **34**:330-339.
130. **Rouf, M. A., e J. L. Stokes.** 1962. Isolation and identification of the sudanophilic granules of *Sphaerotilus natans*. *Journal of Bacteriology* **83**:343-347.

131. **Ruoff, K. L., R. A. Whiley, e D. Beighton.** *Streptococcus*, p. 283-296. In E. J. B. P. R. Murray, M. A. Pfaller, F. C. Tenover, e R. H. Tenover (ed.), Manual of Clinical Microbiology, 7 ed.
132. **Russell, F. M., J. R. Carapetis, S. Ketawai, V. Kunabuli, M. Taoi, S. Biribo, A. Seduadua, e E. K. Mulholland.** 2006. Pneumococcal nasopharyngeal carriage and patterns of penicillin resistance in young children in Fiji. *Annals of Tropical Paediatrics* **26**:187-197.
133. **Sa-Leao, R., S. E. Vilhelmsson, H. de Lencastre, K. G. Kristinsson, e A. Tomasz.** 2002. Diversity of penicillin-nonsusceptible *Streptococcus pneumoniae* circulating in Iceland after the introduction of penicillin-resistant clone Spain(6B)-2. *Journal of Infectious Diseases* **186**:966-975.
134. **Sadowy, E., R. Izdebski, A. Skoczynska, P. Grzesiowski, M. Gniadkowski, e W. Hryniewicz.** 2007. Phenotypic and molecular analysis of penicillin-nonsusceptible *Streptococcus pneumoniae* isolates in Poland. *Antimicrobial Agents Chemotherapy* **51**:40-47.
135. **Saha, S. K., N. Rikitomi, M. Ruhulamin, H. Masaki, M. Hanif, M. Islam, K. Watanabe, K. Ahmed, K. Matsumoto, R. B. Sack, e T. Nagatake.** 1999. Antimicrobial resistance and serotype distribution of *Streptococcus pneumoniae* strains causing childhood infections in Bangladesh, 1993 to 1997. *Journal of Clinical Microbiology* **37**:798-800.
136. **Sahu, R. K., S. Mordechai, S. Pesakhov, R. Dagan, e N. Porat.** 2006. Use of FTIR spectroscopy to distinguish between capsular types and capsular quantities in *Streptococcus pneumoniae*. *Biopolymers* **83**:434-442.

137. **Sandt, C., C. Madoulet, A. Kohler, P. Allouch, C. De Champs, M. Manfait, e G. D. Sockalingum.** 2006. FT-IR microspectroscopy for early identification of some clinically relevant pathogens. *Journal of Applied Microbiology* **101**:785-797.
138. **Savitzky, A., e M. J. E. Golay.** 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **36**:1627-1633.
139. **Schmalreck, A. F., P. Tränkle, E. Vanca, e R. Blaschke-Hellmessen.** 1998. Differentiation and characterization of yeasts pathogenic for humans (*Candida albicans*, *Exophiala dermatidis*) and algae pathogenic for animals (*Prototheca spp.*) using Fourier transform infrared spectroscopy (FTIR) in comparison with conventional methods. *Mycoses* **41**:71-77.
140. **Schmitt, J., T. Udelhoven, D. Naumann, e H. C. Flemming.** 1998. Stacked spectra in biomedical applicationspectral data processing and artificial neural networks applies to FT-IR and FT-Raman Proceedings in SPIE International Society for Optical Engineering **3257**:236-244.
141. **Schuster, K. C., F. Mertens, e J. R. Gapes** 1999. FTIR spectroscopy applied to bacterial cells as a novel method for monitoring complex biotechnological processes. *Vibrational Spectroscopy* **19**:467-477.
142. **Seltmann, G., W. Voigt, e W. Beer.** 1994. Application of physico-chemical typing methods for the epidemiological analysis of *Salmonella enteritidis* strains of phage type 25/17. *Epidemiology and Infection* **113**:411-424.

143. **Sener, B., L. McGee, A. Pinar, e O. Eser.** 2006. Genomic backgrounds of drug-resistant *Streptococcus pneumoniae* in Ankara, Turkey: Identification of emerging new clones. *Microbial Drug Resistance* **12**:109-114.
144. **Serrano, I., J. Melo-Cristino, J. A. Carriço e M. Ramirez.** 2005. Characterization of the genetic lineages responsible for pneumococcal invasive disease in Portugal. *Journal of Clinical Microbiology* **43**:1706-1715.
145. **Serrano, I., M. Ramirez, Grupo português de vigia para o estudo de patógenos respiratórios (Portuguese surveillance group for the study of respiratory pathogens) e J. Melo-Cristino.** 2004. Invasive *Streptococcus pneumoniae* from Portugal: implications for vaccination and antimicrobial therapy. *Clinical Microbiology and Infection* **10**:652-656.
146. **Shaps, R. H., e J. F. Sprouse.** 1980. Automated infrared spectra search and display. *European Spectroscopy News* **32**:39-42.
147. **Smith, A. J., J. Jefferies, S. C. Clarke, C. Dowson, G. F. Edwards, e T. J. Mitchell.** 2006. Distribution of epidemic antibiotic-resistant pneumococcal clones in Scottish pneumococcal isolates analysed by multilocus sequence typing. *Microbiology* **152**:361-365.
148. **Smith, A. M., e K. P. Klugman.** 2005. Amino acid mutations essential to production of an altered PBP 2X conferring high-level β -lactam resistance in clinical isolate of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy* **49**:4622-4627.

149. **Smith, B. C.** 1996. Fundamentals of Fourier transform infrared spectroscopy. CCR press LLC, Florida.
150. **Sockalingum, G. D., Bouhedja, W., P. Pina, P. Allouch, C. Mandray, R. Lábia, J. M. Millot, e M. Manfait.** 1997. ATR-FTIR Spectroscopic investigation of imipenem susceptible and -resistant *Pseudomonas aeruginosa* isogenic strains. Biochemical and Biophysical Research Communications **232**:240-246.
151. **Sockalingum, G. D., W. Bouhedja, P. Pina, P. Allouch, C. Bloy, e M. Manfait.** 1998. FT-IR spectroscopy as an emerging method for rapid characterization of microorganisms. Cellular and Molecular Biology **44**:261-269.
152. **Sogstad, M. K., E. A. Hoiby, e D. A. Caugant.** 2006. Molecular characterization of non-penicillin-susceptible *Streptococcus pneumoniae* in Norway. Journal of Clinical Microbiology **44**:3225-3230.
153. **Song, J. H., N. Y. Lee, S. Ichiyama, R. Yoshida, Y. Hirakata, W. Fu, A. Chongthaleong, N. Aswapokee, C. H. Chiu, M. K. Lalitha, K. Thomas, J. Perera, T. T. Yee, F. Jamal, U. C. Warsa, B. X. Vinh, M. R. Jacobs, P. C. Appelbaum, e C. H. Pai.** 1999. Spread of drug-resistant *Streptococcus pneumoniae* in Asian countries: Asian Network for Surveillance of Resistant Pathogens (ANSORP) Study. Clinical Infectious Diseases **28**:1206-1211.
154. **Song, J. H., S. I. Jung, K. S. Ko, N. Y. Kim, J. S. Son, H. H. Chang, H. K. Ki, W. S. Oh, J. Y. Suh, K. R. Peck, N. Y. Lee, Y. Yang, Q. Lu, A. Chongthaleong, C. H. Chiu, M. K. Lalitha, J. Perera, T. T. Yee, G. Kumarasinghe, F. Jamal, A. Kamarulzaman, N. Parasakthi, P. H.**

- Van, C. Carlos, T. So, T. K. Ng, e A. Shibl.** 2004. High prevalence of antimicrobial resistance among clinical *Streptococcus pneumoniae* isolates in Asia (an ANSORP study). *Antimicrobial Agents Chemotherapy* **48**:2101-2107.
155. **Sorensen, U. B.** 1993. Typing of pneumococci by using 12 pooled antisera. *Journal of Clinical Microbiology* **31**:2097-2100.
156. **Stuart, B.** 2004. *Infrared spectroscopy: Fundamentals and applications*, Chichester, Inglaterra.
157. **Tamm, E., P. Naaber, M. Maimets, M. Oona, S. Koljalg, e I. Lutsar.** 2007. Antimicrobial susceptibility and serogroup/serotype distribution of nasopharyngeal isolates of *Streptococcus pneumoniae* in healthy Estonian children in 1999-2003. *Clinical Microbiology and Infection* **13**:24-826.
158. **Tarasi, A., N. Sterk-Kuzmanovic, K. Sieradzki, S. Schoenwald, R. Austrian, e A. Tomasz.** 1995. Penicillin-resistant and multidrug-resistant *Streptococcus pneumoniae* in a pediatric hospital in Zagreb, Croatia. *Microbial Drug Resistance* **1**:169-176.
159. **Timmins, E. M., D. E. Quain, e R. Goodacre.** 1998. Differentiation of brewing yeast strains by pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. *Yeast* **14**:885-893.
160. **Timmins, E. M., S. A. Howell, B. K. Alsberg, W. C. Noble, e R. Goodacre.** 1998. Rapid differentiation of closely related *Candida* species and strains by pyrolysis-mass spectrometry and Fourier transform-infrared spectroscopy. *Journal of Clinical Microbiology* **36**:367-374.

161. **Tintelnot, K., G. Haase, M. Seibold, F. Bergmann, M. Staemmler, T. Franz, e D. Naumann.** 2000. Evaluation of phenotypic markers for selection and identification of *Candida dubliniensis*. *Journal of Clinical Microbiology* **38**:1599-1608.
162. **Trzciński, K., C. M. Thompson, e M. Lipsitch.** 2004. Single-step capsular transformation and acquisition of penicillin resistance in *Streptococcus pneumoniae*. *Journal of Bacteriology* **186**:3447-3452.
163. **Udelhoven, T., D. Naumann, e J. Schmitt.** 2000. Development of a hierarchical classification systems with artificial neural networks and FT-IR spectra for the identification of bacteria. *Applied Spectroscopy* **54**:1471-1479.
164. **Valles, X., B. Flannery, A. Roca, I. Mandomando, B. Sigauque, S. Sanz, A. Schuchat, M. Levine, M. Soriano-Gabarro, e P. Alonso.** 2006. Serotype distribution and antibiotic susceptibility of invasive and nasopharyngeal isolates of *Streptococcus pneumoniae* among children in rural Mozambique. *Tropical Medicine and International Health* **11**:358-366.
165. **Van Der Mei, H. C., D. Naumann, e H. J. Busscher.** 1993. Grouping of oral streptococcal species using Fourier-transform infrared spectroscopy in comparison with classical microbiological identification. *Archives of Oral Biology* **38**:1013-1019.
166. **Van der Mei, H. C., D. Naumann, e H. J. Busscher.** 1996. Grouping of *Streptococcus mitis* strains grown on different growth media by FT-IR. *Infrared Physics & Technology* **37**:561-564.

167. **Vanhoof, R., M. Carpentier, R. Cartuyvels, S. Damee, O. Fagnart, J. Frans, B. Gordts, Y. Glupczynski, P. Goffinet, D. Govaerts, P. Lefevre, M. Lontie, K. Magerman, I. Mans, F. Meunierz, F. Moonens, I. Surmont, E. Van Bossuyt, M. Van De Vyvere, J. Van Eldere, H. Van Landuyt, L. Van Nimmen, R. Van Noyen, e J. Verhaegen.** 2006. Surveillance of antibiotic resistance in non invasive clinical isolates of *Streptococcus pneumoniae* collected in Belgium during winters 2003 and 2004. *Acta Clinica Belgica* **61**:49-57.
168. **Winder, C. L., e R. Goodacre.** 2004. Comparison of didduse-reflectance absorbance and attenuated total reflectance FT-IR for the discremination of bacteria. *The Analyst* **129**:1118-1122.
169. **Winder, C. L., S. V. Gordon, J. Dale, R. Glyn Hewinson, e R. Goodacre.** 2006. Metabolic fingerprints of *Mycobacterium bovis* cluster with molecular type: implications for genotype-phenotype links. *Microbiology* **152**:2757-2765.
170. **Wise, B. M., N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig, e R. S. Koch.** 2004. PLS_Toolbox 3.5 for use with MATLAB™. Eigenvector Research Incorporated, Manson, USA.
171. **Wold, S.** 1976. Pattern recognition by means of disjiont principal component models. *Pattern Recognition* **8**:127-139.
172. **World Health Organization.** 2003. Immunization, vaccines and biologicals. World Health Organization, Geneve, Suíça.
173. **Yu, C., e J. Irudayaraj.** 2005. Spectroscopic charecterization of microorganisms by Fourier transform infrared microspectroscopy. *Biopolymers* **77**:368-377.

174. **Yu, V. L., C. C. Chiou, C. Feldman, A. Ortqvist, J. Rello, A. J. Morria, L. M. Baddour, C. M. Luna, D. R. Snyderman, M. Ip, W. C. Ko, M. B. Chedid, A. Andremon, e K. P. Klugman.** 2003. An international prospective study of pneumococcal bacteremia: correlation with in vitro resistance, antibiotics administered, and clinical outcome. *Clinical Infectious Diseases* **37**:230-237.
175. **Zdenek, F., e S. Hermann.** 2001. An attempt to differentiate *Pseudomonas spp.* and other soil bacteria by FT-IR spectroscopy. *European Journal of Soil Biology* **37**:137-143.
176. **Zemlickova, H., O. Melter, e P. Urbaskova.** 2006. Epidemiological relationships among penicillin non-susceptible *Streptococcus pneumoniae* strains recovered in the Czech Republic. *Journal of Medical Microbiology* **55**:437-442.
177. **Zeroual, W., M. Manfait, e C. Choisy.** 1995. FT-IR spectroscopy study of perturbations induced by antibiotic on bacteria (*Escherichia coli*). *Pathologie Biologie* **43**:300-305.
178. **Zhanel, G. G., X. Wang, K. Nichol, A. Nikulin, A. K. Wierzbowski, M. Mulvey, e D. J. Hoban.** 2006. Molecular characterisation of Canadian paediatric multidrug-resistant *Streptococcus pneumoniae* from 1998-2004. *International Journal of Antimicrobial Agents* **28**:465-471.

VII. ANEXOS

ANEXO I

Composição dos meios e soluções

Composição do meio de cultura “Gelose de Sangue” da Biomérieux, ref^a 43001

Composição do reagente de saponificação (Reagente GC #1) para a extracção dos ácidos gordos

Metanol (HPLC grade)	150 ml
Água desionizada	150 ml
Hidróxido de Sódio (ACS)	45g

Composição do reagente de Metilação (reagente GC #2) para a extracção dos ácidos gordos

Metanol (HPLC grade)	275 ml
Ácido Clorídrico 6,00N	325 ml

Composição do reagente de Extracção (reagente GC #3) para extracção dos ácidos gordos

Hexano (HPLC grade)	200 ml
MTBE (HPLC grade)	200 mL

Composição do reagente de Lavagem (reagente GC #4) para a extracção dos ácidos gordos

Água Desionizada	900 ml
Hidróxido de Sódio (ACS)	10,8 g

Composição da Solução saturada de Cloreto de sódio (reagente GC #5) para extracção dos ácidos gordos

Água Desionizada	100 ml
Cloreto de sódio (ACS)	40 g

ANEXO II

Explicação pormenorizada dos parâmetros usados no teste de Qualidade

A avaliação do teste de qualidade é feita com base na escolha de números de onda característicos de determinadas bandas, que devem estar presentes num espectro de um microrganismo, e de números de onda nos quais não se devem observar bandas, quando se trata de analisar um microrganismo. Assim a explicação do conteúdo da Tabela II.2 é a seguinte ⁽²⁷⁾:

- Amplitude x – 1: Os valores de absorvância máximo e mínimo do espectro, que estejam no intervalo do número de onda definido na primeira coluna da tabela, serão determinados. Depois, será calculada a diferença entre os valores de absorvância determinados. Esta diferença deverá ser superior a 0,2 (mínimo delta y) e inferior a 2 (máximo delta y).
- Amplitude x – 4: Nesta zona não devem existir bandas para as bactérias, assim será escolhida para a determinação do ruído. Para isso, é calculada a primeira derivada e são determinados os valores de absorvância máximos e mínimos neste intervalo ($2000\text{-}2100\text{ cm}^{-1}$). A diferença entre estes valores não deverá ser superior ao valor de ruído máximo definido na tabela (0,00015).
- Amplitude x – 5: Para avaliar a quantidade de vapor de água presente na amostra, deve-se avaliar a banda da água numa zona em que não existam bandas da amostra. Esta zona é entre $1847\text{ e }1837\text{ cm}^{-1}$. Deve ser calculada a primeira derivada e determinados os valores de absorvância máximos e mínimos para este intervalo de números de onda. A diferença entre as absorvâncias determinadas não deverá ser superior ao valor definido na tabela para “água máx”.

- Amplitude x -2: Para avaliar a razão sinal/ruído deve escolher-se uma zona do espectro que contenha uma banda característica, para se ter um valor para o sinal, e uma zona que não tenha bandas, para se avaliar o ruído. Todas as bactérias têm uma banda característica que corresponde à Amida I ($1700\text{-}1600\text{ cm}^{-1}$). Como já foi referido, entre os 2100 e 2000 cm^{-1} , não existe nenhuma descrição, até à data, de alguma banda, quando se estudam bactérias. Assim, no teste de qualidade é determinada a primeira derivada do espectro e, são determinados os valores máximos e mínimos de absorvância para o intervalo $1700\text{-}1600\text{ cm}^{-1}$. A diferença entre estes valores gera um S_1 , um parâmetro que corresponde, aproximadamente, à intensidade da banda definida.

Este S_1 é usado para determinar a razão sinal/ruído, que se calcula dividindo S_1 pelo ruído. Esta razão deverá ser superior a 150.

Dividindo S_1 pelo sinal do vapor de água, a razão deverá ser superior a 100.

- Amplitude x - 3: Nesta gama de comprimentos de onda encontra-se uma banda correspondente à vibração dos anéis de açúcar, também presentes nas bactérias. Nesta gama estuda-se uma segunda razão sinal/ruído. Assim, determina-se a primeira derivada e os máximos e mínimos de absorvância, para este intervalo. Depois, calcula-se a diferença entre estas absorvâncias obtendo-se um valor para S_2 . Dividindo este S_2 pelo valor determinado para o ruído, dever-se-á obter um valor superior ao definido em sinal/ruído. Dividindo S_2 pelo valor calculado de vapor de água (em “água máx”), obtém-se um valor que deverá ser superior ao valor tabelado para o parâmetro sinal/água.

- Amplitude $\times - 6$: Nesta gama de números de onda, não deverão existir bandas, ou seja, permite determinar as interferências. A diferença entre o máximo e mínimo de absorvância da primeira derivada, na zona entre $2200-2000\text{ cm}^{-1}$, deverá ser superior ao definido na tabela, na linha correspondente a interferências.

ANEXO III

Tabela VII.1 – Resultados para a análise de ácidos gordos por GC das estirpes de *S. pneumoniae*. Para as estirpes que foram analisadas em dois dias, a estirpe do segundo dia tem a letra “a” após a indicação do serótipo. As linhas sublinhadas a cinzento correspondem a estirpes que tiveram de ser concentradas numa linha de azoto de modo a que a resposta total do GC fosse superior a 50000, resposta mínima exigida pela técnica.

Ácido gordo designado	12:0	13:0	14:0	13:0 2OH	15:1 w5c	15:0	16:1w9c	Sum in feature 3	16:1 w5c	16:0	17:1iso w5c	17:1 anteiso w9c	17:1 w8c	17:0	Sum in feature 6	18:1w9c	18:1 w7c	18:1 w5c	18:0	19:1 iso l	19:0 iso	20:4 w6,9,12,15c
Tempo de retenção	1,648	1,91	2,2	2,268	2,482	2,508	2,761	2,774	2,802	2,825	2,98	3,002	3,084	3,144	3,383	3,395	3,411	3,439	3,46	3,614	3,655	3,911
1999V0490\$su19A	8,74	0,53	17,72	0,00	0,00	1,12	2,08	11,06	0,99	32,08	0,00	0,00	0,00	1,04	5,08	5,16	6,55	0,00	7,19	0,00	0,00	0,66
1999V0772\$su9V	13,42	0,60	19,83	0,00	0,00	1,11	1,09	5,45	0,57	32,27	0,00	0,00	0,00	1,11	4,78	4,71	5,38	0,00	9,08	0,00	0,00	0,60
1999V0975\$su23F	8,16	0,33	15,46	0,00	0,00	0,76	1,61	8,56	0,86	36,29	0,00	0,00	0,40	1,09	4,94	5,11	7,69	0,38	7,74	0,00	0,00	0,63
1999V0983\$su14	10,81	0,60	21,25	0,00	0,00	1,19	1,49	8,20	0,77	35,36	0,00	0,00	0,00	0,88	4,01	3,93	5,27	0,00	5,73	0,00	0,00	0,50
1999V0993\$su9V	7,42	0,46	16,23	0,00	0,00	1,10	1,66	9,44	0,92	35,12	0,00	0,00	0,46	1,17	4,92	5,59	7,30	0,33	7,21	0,00	0,00	0,68
1999V1010\$su19Aa	7,40	0,51	15,61	0,00	0,85	1,13	1,91	8,46	0,83	33,22	0,00	0,00	0,58	1,41	4,50	5,52	8,10	0,00	9,06	0,00	0,44	0,49
2000V0193\$su19A	8,79	0,00	15,89	0,00	0,00	1,13	1,77	10,23	0,91	32,83	0,00	0,00	0,00	1,31	5,24	6,09	6,65	0,00	8,54	0,00	0,00	0,61
2000V0324\$su19A	8,55	0,54	17,12	0,00	0,00	1,12	1,85	9,23	0,84	33,74	0,00	0,00	0,48	1,17	4,70	5,61	6,39	0,00	7,62	0,00	0,48	0,55
2000V0324\$su19Aa	9,39	0,00	18,28	0,00	0,00	0,70	1,64	8,44	0,76	33,09	0,91	0,00	0,00	0,91	4,80	6,88	5,78	0,00	8,67	0,00	0,00	0,67
2000V0906\$suNT	6,82	0,00	14,46	0,00	0,00	1,30	1,69	8,07	0,82	33,81	0,00	0,00	0,00	1,47	5,79	7,11	7,92	0,00	10,03	0,00	0,00	0,70
2000V0906\$suNTa	6,89	0,00	14,27	0,00	0,00	0,84	1,57	7,73	0,77	31,85	0,00	0,00	0,00	1,20	6,25	9,19	7,50	0,00	11,14	0,00	0,00	0,80
2000V1210\$su14	7,62	0,40	15,54	0,00	0,00	0,82	1,54	8,58	0,87	35,67	0,00	0,00	0,37	1,09	3,86	4,73	8,81	0,45	7,84	0,40	0,59	0,83
2000V1217\$su23F	7,96	0,00	15,70	0,00	0,00	0,94	1,62	8,67	0,83	35,13	0,00	0,00	0,00	1,30	5,15	5,99	7,90	0,00	7,82	0,00	0,00	1,01
2000V1217\$su23Fa	9,63	0,00	17,49	0,00	0,00	0,66	1,52	7,80	0,75	33,73	0,00	0,00	0,00	0,88	5,01	7,10	5,86	0,00	8,77	0,00	0,00	0,80
2000V1226\$su9V	8,80	0,00	17,20	0,00	0,00	1,18	1,53	8,14	0,75	34,50	0,00	0,00	0,00	1,34	5,36	6,03	6,96	0,00	8,19	0,00	0,00	0,00
2000V1226\$su9Va	10,02	0,00	17,60	0,00	0,00	0,72	1,44	7,45	0,73	32,66	0,00	0,00	0,00	0,93	5,31	7,29	5,91	0,00	9,22	0,00	0,00	0,71
2001V0172\$su9V	9,54	0,00	19,43	0,00	0,00	1,28	1,64	8,88	0,80	34,36	0,00	0,00	0,00	1,11	5,06	5,13	5,85	0,00	6,93	0,00	0,00	0,00
2001V0172\$su9Va	9,65	0,00	18,46	0,00	0,00	0,73	1,44	7,81	0,72	33,54	0,00	0,00	0,00	0,93	5,09	6,84	5,52	0,00	8,59	0,00	0,00	0,70
2001V0180\$su9V	9,27	0,68	19,32	0,00	0,00	1,29	1,67	9,11	0,83	34,84	0,00	0,00	0,00	1,20	4,27	4,63	5,62	0,00	6,75	0,00	0,00	0,52
2001V0462\$su14	8,23	0,00	17,63	0,00	0,00	1,13	1,45	7,54	0,72	35,08	0,00	0,00	0,00	1,31	5,22	5,94	6,42	0,00	9,33	0,00	0,00	0,00
2001V0761\$su14	9,29	0,00	18,06	0,00	0,00	1,13	1,47	7,95	0,75	34,69	0,00	0,00	0,00	1,18	4,75	5,23	6,27	0,00	9,24	0,00	0,00	0,00
2001V0919\$su23F	7,38	0,48	15,16	0,00	0,50	1,04	2,01	10,06	0,98	31,81	0,65	0,00	0,60	1,21	4,50	5,62	8,65	0,43	7,12	0,42	0,76	0,62
2001V0919\$su23Fa	7,88	0,00	15,32	0,00	0,00	0,66	1,64	9,02	0,84	33,72	0,00	0,00	0,00	1,03	4,39	7,00	8,27	0,00	9,69	0,00	0,00	0,54

Tabela VII.1 (continuação)

Ácido gordo designado	12:0	13:0	14:0	13:0 2OH	15:1 w5c	15:0	16:1w9c	Sum in feature 3	16:1 w5c	16:0	17:1iso w5c	17:1 anteiso w9c	17:1 w8c	17:0	Sum in feature 6	18:1w9c	18:1 w7c	18:1 w5c	18:0	19:1 iso l	19:0 iso	20:4 w6,9,12,15c
Tempo de retenção	1,648	1,91	2,2	2,268	2,482	2,508	2,761	2,774	2,802	2,825	2,98	3,002	3,084	3,144	3,383	3,395	3,411	3,439	3,46	3,614	3,655	3,911
2001V1072\$su23F	7,46	0,42	14,05	0,00	0,00	0,87	1,52	8,21	0,87	36,89	0,00	0,00	0,49	1,39	4,08	4,89	9,12	0,43	8,71	0,00	0,00	0,60
2001V1096\$su14	8,27	0,64	16,82	0,00	0,00	1,24	1,43	8,03	0,76	33,54	0,00	0,00	0,44	1,37	4,78	6,08	6,96	0,00	8,58	0,00	0,50	0,55
2001V1123\$su23F	10,06	0,54	18,12	0,00	0,00	0,96	1,44	7,61	0,70	35,35	0,00	0,00	0,00	1,13	4,90	5,16	6,02	0,00	7,12	0,00	0,00	0,90
2001V1152\$su14	8,93	0,00	16,56	0,00	0,00	1,15	1,45	8,22	0,75	33,59	0,00	0,00	0,00	1,37	5,24	6,13	7,02	0,00	9,57	0,00	0,00	0,00
2001V1152\$su14a	9,17	0,00	16,82	0,00	0,00	0,67	1,42	7,32	0,72	32,85	0,00	0,00	0,00	0,94	5,57	7,54	6,15	0,00	10,00	0,00	0,00	0,83
2001V1222\$su19A	7,51	0,00	16,57	0,00	0,00	0,87	1,72	9,80	0,93	36,37	0,00	0,00	0,00	1,04	3,97	4,54	8,33	0,00	7,86	0,00	0,00	0,48
2001V1222\$su19Aa	7,69	0,00	16,47	0,00	0,00	0,66	1,54	8,62	0,85	35,26	0,00	0,00	0,00	0,95	4,36	6,59	7,20	0,39	8,84	0,00	0,00	0,60
2001V1277\$su9V	9,86	0,55	19,58	0,00	0,00	1,09	1,49	7,73	0,71	34,67	0,00	0,00	0,00	1,23	4,52	4,75	5,74	0,00	7,56	0,00	0,00	0,52
2001V1277\$su9Va	10,30	0,00	19,09	0,00	0,00	0,67	1,25	6,33	0,63	33,07	0,00	0,00	0,00	0,90	5,47	7,32	5,11	0,00	9,24	0,00	0,00	0,61
2002V0067\$su23F	8,34	0,00	16,45	0,00	0,00	1,08	1,81	9,32	0,88	36,53	0,00	0,00	0,00	1,18	4,56	5,08	6,88	0,00	7,35	0,00	0,00	0,54
2002V0067\$su23Fa	7,56	0,00	15,24	0,00	0,00	0,62	1,69	8,72	0,89	34,55	0,00	0,00	0,00	1,00	5,15	7,32	7,32	0,00	9,25	0,00	0,00	0,67
2002V0488\$su14	6,33	0,47	14,43	0,00	0,76	1,05	2,16	10,93	1,12	31,88	0,00	0,00	0,58	1,27	5,11	6,20	8,65	0,00	8,09	0,00	0,55	0,44
2002V0701\$su19	10,05	0,68	18,84	0,00	0,00	1,22	1,81	9,44	0,90	33,09	0,00	0,00	0,00	1,10	4,05	4,60	6,91	0,00	6,85	0,00	0,00	0,47
2002V0929\$su14	6,51	0,48	12,90	0,00	0,00	0,96	1,49	8,65	0,88	38,11	0,00	0,00	0,46	1,48	4,55	5,45	8,47	0,00	9,05	0,00	0,00	0,57
2002V0929\$su14a	6,31	0,00	11,77	0,00	0,00	0,58	1,52	8,65	0,85	37,06	0,00	0,00	0,00	1,12	5,13	7,49	7,72	0,41	10,63	0,00	0,00	0,76
2002V1462\$su14	9,38	0,51	16,04	0,00	0,00	0,91	1,20	5,92	0,60	37,31	0,00	0,00	0,00	1,37	4,35	4,70	7,32	0,00	8,96	0,00	0,00	0,64
2002V1463\$suNT	7,79	0,00	16,66	0,00	0,00	0,92	2,06	9,68	0,95	32,57	0,00	0,00	0,00	1,17	5,75	6,34	7,03	0,00	8,37	0,00	0,00	0,70
2002V1463\$suNTa	7,21	0,00	14,96	0,00	0,00	0,66	1,69	7,96	0,81	31,47	0,00	0,00	0,00	1,07	6,25	9,21	6,69	0,00	11,04	0,00	0,00	0,97
2003V0033\$su19A	8,19	0,62	15,39	0,00	0,00	1,24	1,78	8,48	0,86	34,58	0,00	0,00	0,58	1,37	4,56	5,61	8,13	0,00	7,99	0,00	0,00	0,61
2003V0033\$su19Aa	8,35	0,33	16,74	0,00	0,00	0,79	1,63	7,94	0,81	33,93	0,00	0,32	0,33	0,95	4,83	7,00	6,18	0,29	8,90	0,00	0,00	0,68
2003V0063\$su23F	7,09	0,45	14,48	0,00	0,00	1,02	1,77	9,19	0,94	33,77	0,00	0,00	0,53	1,29	5,50	6,16	7,91	0,00	8,66	0,00	0,60	0,64
2003V0063\$su23Fa	7,15	0,00	14,05	0,00	0,00	0,80	1,45	8,28	0,84	32,69	0,00	0,00	0,00	1,12	5,67	8,49	8,11	0,00	10,59	0,00	0,00	0,75
2003V0076\$su19A	9,02	0,70	18,16	0,00	0,00	1,34	1,98	11,36	1,02	33,22	0,00	0,00	0,52	1,12	3,80	4,35	6,92	0,00	6,48	0,00	0,00	0,00
2003V0076\$su19Aa	7,79	0,00	15,18	0,00	0,00	0,81	1,62	9,61	0,91	33,40	0,00	0,00	0,00	1,06	4,67	7,08	7,71	0,00	9,39	0,00	0,00	0,78
2003V0088\$su23F	9,17	0,00	17,18	0,00	0,00	0,95	1,72	8,87	0,87	36,04	0,00	0,00	0,00	1,14	4,65	5,02	6,55	0,00	7,31	0,00	0,00	0,53
2003V0088\$su23Fa	7,30	0,00	14,27	0,00	0,00	0,59	1,45	8,18	0,81	33,74	0,00	0,00	0,00	1,06	5,16	7,82	8,05	0,44	10,34	0,00	0,00	0,78
2003V0148\$su23F	8,96	0,72	18,08	0,00	0,00	1,43	2,03	11,07	0,93	31,36	0,00	0,00	0,55	1,10	5,04	5,28	5,85	0,00	6,97	0,00	0,00	0,62
2003V0262\$su9V	8,64	0,57	17,57	0,00	0,00	1,14	1,49	8,36	0,76	33,33	0,00	0,00	0,00	1,31	4,96	6,08	6,91	0,00	8,10	0,00	0,00	0,76

Tabela VII.1 (continuação)

Ácido gordo designado	12:0	13:0	14:0	13:0 2OH	15:1 w5c	15:0	16:1w9c	Sum in feature 3	16:1 w5c	16:0	17:1iso w5c	17:1 anteiso w9c	17:1 w8c	17:0	Sum in feature 6	18:1w9c	18:1 w7c	18:1 w5c	18:0	19:1 iso l	19:0 iso	20:4 w6,9,12,15c
Tempo de retenção	1,648	1,91	2,2	2,268	2,482	2,508	2,761	2,774	2,802	2,825	2,98	3,002	3,084	3,144	3,383	3,395	3,411	3,439	3,46	3,614	3,655	3,911
2003V0319\$suNT	7,66	0,00	16,51	0,00	0,00	0,98	1,50	7,80	0,82	36,71	0,00	0,00	0,00	1,35	4,79	4,99	7,53	0,00	8,37	0,00	0,00	0,98
2003V0416\$suNT	7,92	0,72	16,26	0,00	0,00	1,30	1,28	6,23	0,66	36,24	0,00	0,00	0,00	1,53	5,32	5,78	6,46	0,00	9,68	0,00	0,00	0,62
2003V0577\$su19A	9,11	0,00	16,05	0,00	0,00	0,92	1,96	8,35	0,99	34,38	0,00	0,00	0,00	1,21	5,22	5,93	6,89	0,00	9,00	0,00	0,00	0,00
2003V0577\$su19Aa	11,41	0,00	19,56	0,00	0,00	0,68	1,66	8,25	0,79	31,56	0,00	0,00	0,00	0,78	4,98	6,67	4,92	0,00	8,12	0,00	0,00	0,62
2003V0583\$suNT	7,56	0,71	17,11	0,00	0,00	1,55	1,65	7,44	0,86	35,18	0,00	0,00	0,00	1,36	5,38	5,54	6,27	0,00	8,77	0,00	0,00	0,62
2003V0583\$suNTa	7,67	0,00	16,51	0,00	0,00	0,97	1,44	6,95	0,79	32,97	0,00	0,00	0,00	1,03	6,62	8,50	6,01	0,00	9,74	0,00	0,00	0,79
2003V0585\$su19A	11,54	0,53	20,65	0,00	0,00	1,12	1,60	8,44	0,77	34,03	0,00	0,00	0,00	1,02	4,36	4,38	4,98	0,00	6,58	0,00	0,00	0,00
2003V0585\$su19Aa	12,68	0,00	21,44	0,00	0,00	0,69	1,49	7,49	0,68	31,01	0,00	0,00	0,00	0,73	4,71	6,24	4,35	0,00	7,81	0,00	0,00	0,68
2003V0785\$su9V	9,17	0,72	18,86	0,00	0,00	1,42	1,78	9,81	0,86	33,03	0,00	0,00	0,52	1,27	4,61	5,28	5,86	0,00	6,17	0,00	0,00	0,64
2003V0785\$su9Va	8,47	0,00	17,29	0,00	0,00	0,83	1,43	8,36	0,80	33,84	0,00	0,00	0,00	1,20	4,74	6,81	6,50	0,00	9,14	0,00	0,00	0,59
2003V0787\$suNT	10,77	0,79	22,05	0,00	0,00	1,42	0,69	2,77	0,00	36,65	0,00	0,00	0,00	1,12	5,74	6,82	3,46	0,00	7,71	0,00	0,00	0,00
2003V0787\$suNTa	11,03	0,48	19,81	0,00	0,00	1,03	0,63	2,39	0,00	32,82	0,00	0,41	0,00	0,99	6,48	9,66	3,29	0,00	10,19	0,00	0,00	0,80
2003V0790\$su14	7,26	0,51	14,59	0,00	0,00	0,96	1,72	8,64	0,90	35,19	0,00	0,00	0,49	1,33	4,32	5,06	9,34	0,49	8,59	0,00	0,00	0,61
2003V0803\$su14	9,73	0,65	19,29	0,00	0,00	1,16	1,67	9,10	0,77	33,20	0,00	0,00	0,00	1,13	4,61	5,14	5,92	0,00	7,13	0,00	0,00	0,51
2003V0803\$su14a	11,54	0,00	20,58	0,00	0,00	0,66	1,53	7,47	0,67	32,34	0,00	0,00	0,00	0,82	4,30	6,11	5,26	0,00	8,14	0,00	0,00	0,58
2003V0813\$su14	9,01	0,00	17,39	0,00	0,00	0,82	1,57	8,38	0,83	36,62	0,00	0,00	0,00	1,05	4,01	4,67	7,49	0,00	8,16	0,00	0,00	0,00
2003V0813\$su14a	7,79	0,00	15,35	0,00	0,00	0,54	1,43	7,57	0,77	35,50	0,00	0,00	0,00	0,79	4,38	6,97	8,79	0,00	9,61	0,00	0,00	0,51
2003V1104\$su14	9,79	0,00	18,87	0,00	0,00	0,91	1,58	7,98	0,79	33,56	0,00	0,00	0,00	0,96	4,90	4,76	6,26	0,00	8,91	0,00	0,00	0,73
2003V1112\$su19A	8,88	0,00	17,39	0,00	0,58	0,77	1,71	9,15	0,88	35,32	0,00	0,00	0,00	1,01	4,54	4,70	6,91	0,00	7,49	0,00	0,00	0,66
2003V1113\$suNT	5,75	0,00	12,23	0,00	0,00	1,23	1,95	8,91	1,02	34,04	0,00	0,00	0,67	1,42	7,08	7,06	8,30	0,00	9,44	0,00	0,00	9,44
2003V1113\$suNTa	7,15	0,00	13,79	0,00	0,00	0,89	1,73	8,31	0,90	33,32	0,00	0,00	0,00	1,13	6,46	8,40	6,99	0,00	10,10	0,00	0,00	0,83
2003V1199\$su14	8,50	0,50	17,62	0,00	0,00	1,00	1,52	8,56	0,79	36,30	0,00	0,00	0,37	1,14	4,17	4,51	6,75	0,00	7,27	0,00	0,43	0,57
2003V1200\$su14	8,83	0,00	17,84	0,00	0,00	0,81	1,47	7,48	0,80	35,46	0,00	0,00	0,00	1,00	5,17	5,37	6,99	0,00	8,12	0,00	0,00	0,65
2003V1200\$su14a	10,67	0,00	19,21	0,00	0,00	0,79	1,39	6,81	0,73	32,78	0,00	0,00	0,00	1,00	5,63	7,24	4,64	0,00	8,39	0,00	0,00	0,73
2003V1203\$suNT	4,85	0,00	18,21	0,00	0,00	0,63	5,30	4,28	0,80	36,80	0,00	0,55	0,00	0,85	4,93	6,48	7,29	0,00	7,65	0,00	0,00	0,72
2003V1203\$suNTa	5,15	0,00	18,72	0,00	0,00	0,52	4,69	4,17	0,70	35,99	0,00	0,00	0,00	0,81	5,02	8,41	6,34	0,00	8,71	0,00	0,00	0,77
2003V1481\$su23F	8,48	0,00	16,41	0,00	0,00	0,71	1,54	8,07	0,79	35,98	0,00	0,00	0,00	1,02	4,08	6,22	7,22	0,00	8,93	0,00	0,00	0,54
2003V1481\$su23F	8,43	0,44	16,72	0,00	0,00	0,94	1,72	8,57	0,85	36,93	0,00	0,00	0,00	1,17	4,01	4,39	7,67	0,00	7,63	0,00	0,00	0,52
2003V1484\$su23F	8,71	0,42	16,96	0,36	0,00	0,97	2,42	12,21	1,06	31,44	0,48	0,00	0,48	1,04	4,02	4,81	7,32	0,00	6,41	0,51	0,00	0,51

Tabela VII.1 (continuação)

Ácido gordo designado	12:0	13:0	14:0	13:0 2OH	15:1 w5c	15:0	16:1w9c	Sum in feature 3	16:1 w5c	16:0	17:1iso w5c	17:1 anteiso w9c	17:1 w8c	17:0	Sum in feature 6	18:1w9c	18:1 w7c	18:1 w5c	18:0	19:1 iso l	19:0 iso	20:4 w6,9,12,15c
Tempo de retenção	1,648	1,91	2,2	2,268	2,482	2,508	2,761	2,774	2,802	2,825	2,98	3,002	3,084	3,144	3,383	3,395	3,411	3,439	3,46	3,614	3,655	3,911
2003V1485\$su19A	11,56	0,59	20,46	0,00	0,00	1,15	1,48	7,50	0,70	35,58	0,00	0,00	0,00	1,05	3,99	4,22	4,90	0,00	6,36	0,00	0,00	0,48
2003V1485\$su19Aa	11,03	0,00	18,74	0,00	0,00	0,81	1,31	7,21	0,66	35,78	0,00	0,00	0,00	0,82	4,33	5,76	5,23	0,00	7,72	0,00	0,00	0,60
2003V1486\$su19A	10,28	0,61	18,79	0,00	0,00	1,17	1,63	9,02	0,84	33,16	0,00	0,00	0,41	1,03	4,40	4,57	6,28	0,34	6,80	0,00	0,00	0,66
2003V1489\$su19A	9,86	0,45	18,30	0,38	0,00	0,91	2,21	11,43	0,96	33,22	0,44	0,00	0,39	0,92	4,08	4,05	5,70	0,00	6,13	0,00	0,00	0,56
2003V1490\$su19A	11,57	0,67	15,99	0,00	0,00	1,28	1,67	7,52	0,75	29,11	0,00	0,00	0,00	1,18	6,70	6,65	5,98	0,00	10,07	0,00	0,00	0,86
2003V1490\$su19Aa	9,15	0,00	16,12	0,00	0,00	0,80	1,69	8,37	0,82	32,54	0,00	0,00	0,00	0,99	5,15	7,14	7,49	0,00	9,19	0,00	0,00	0,55
2003V1494\$su14	9,72	0,61	18,86	0,00	0,00	1,11	1,38	7,81	0,70	34,12	0,00	0,00	0,00	1,26	4,38	5,10	5,99	0,00	8,96	0,00	0,00	0,00
2003V1495\$su19A	11,80	0,00	20,72	0,00	0,00	1,05	2,28	11,64	0,93	31,64	0,00	0,00	0,00	0,92	4,03	4,24	4,67	0,00	6,08	0,00	0,00	0,00
2003V1528\$su9V	9,49	0,00	19,32	0,00	0,00	0,96	1,58	8,57	0,77	35,37	0,00	0,00	0,00	1,00	4,14	4,25	6,29	0,00	7,72	0,00	0,00	0,55
2003V1528\$su9Va	10,68	0,00	19,71	0,00	0,00	0,67	1,51	7,75	0,71	33,84	0,00	0,00	0,00	0,89	4,11	6,05	5,74	0,00	8,33	0,00	0,00	0,00

